THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Présentée et soutenue par :
Nicolas REGIS

Le mardi 9 décembre 2014

Titre :

Toward the real time estimation of the attentional state through ocular activity analysis

---

EDSYS : Systèmes embarqués 4200046

Unité de recherche :

Commande des Systèmes et Dynamique du Vol

Directeur(s) de Thèse :

Catherine TESSIER
Frédéric DEHAIS

Rapporteurs :

Thierry BACCINO
José CAÑAS DELGADO

Autre(s) membre(s) du jury :

Philippe PALANQUE
Michael FEARY

"Keep up the good work !"

Cette thèse est dédiée à mes parents.

REMERCIEMENTS

Ce mémoire de thèse, qui referme la page de mes études scientifiques, me parait être le tout dernier appui avant de prendre mon envol. Et, alors que le sol s'efface doucement, je voudrais remercier du fond du cœur les personnes qui m'ont porté, poussé, soutenu, ramassé, encouragé, encore et toujours...

Avant la thèse, il y a eu Supaero, dont l'encadrement m'a réconcilié avec l'école en général. Je pense notamment à Caroline pour son énergie irrésistible, à Maryse pour sa bonne humeur, à MFP pour son écoute et ses conseils, à Frédéric pour sa bienveillance (!), à Ausias, Stéphane et Daniel pour le show, et à Guy et Xavier, les meilleurs !

Merci à toute l'équipe du CAS, Eric, Dominique, Patrice, Guillaume, Jérôme, Étienne, Jérémie, Daniel, Bruno, Frédéric, Fabrice, Franck. Une pensée toute particulière pour Marie pour ton soutien, Christian pour ta bienveillance, et Philippe pour m'avoir appris à voler !
Puis il y a le groupe d'élite des FH, Thibault, Gautier, Louise, Eve, Benoit, Kevin. Sébastien, merci pour ton franc soutien amigo, on sait tous les deux pourquoi ça roule pour moi aujourd'hui ! Charlie, Ricou, on a même réussi à écrire des articles ensemble ! Vsevolod, Michael, on en fera une de boîte un jour !
Il est dur d'imaginer le soutien nécessaire pour mener une thèse au bout, je l'ai compris au côté de Catherine et Frédéric qui ont toujours réussi à me faire avancer même dans les moments les plus difficiles, où le doute faisait place au désarroi. Frédéric, merci pour tout, du premier jour où l'on s'est rencontré où tu m'avais convaincu en 5 minutes de venir bosser avec toi, jusqu'aux toutes dernières relectures à bout de souffle... avec la saucisse en fil rouge bien sûr !
Je souhaite aussi remercier Michael Feary pour son accueil à la NASA, et Joy Taylor à Stanford. Thank you Michael, you definitely broadend my horizons by supporting my visit both from a material and human point of view. Thank you Joy for giving me the chance to work with your wonderful team. Cet échange n'aurait pas eu lieu sans le soutien généreux d'EDSYS et de la Fondation ISAE-SUPAERO. Merci.

Finalement, je remercie à genou mes amis grâce auxquels j'ai pu déployer mes ailes. Trouver la Chabale en arrivant à Sup fut un cadeau du ciel. Les mots commencent à manquer...

Et mes parents...

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AOI	Area of Interest
BEA	Bureau Enquete Analyse (french safety board)
BU	Bottom-Up
CFIT	Controlled Flight Into Terrain
EOG	Electro-oculography
ET	Eye-Tracker, Eye-Tracking
FEF	Front Eye Field
FIT	Feature Integration Theory
FOA	Focus Of Attention
LC-NE	Locus cœruleus-norepinephrine
LCI	Lost of Control In-flight
LGN	Lateral Geniculate Nuclei
M-	Magnocellular (channels, layers, projections)
MIT	Massachusetts Institute of Technology
MT	Middle Temporal
MST	Middle Superior Temporal
P-	Parvocellular (channels, layers, projections)
P	Pulvinar
PPC	Posterior Parietal Cortex
ROI	Region Of Interest
SA	Situation Awareness
SEEV	Saliency Effort Expectation Value (model of selective attention)
SC	Superior Colliculus
TD	Top-Down
V1	Primary Visual Cortex
VOG	Video-OculoGraphy

1

INTRODUCTION

RÉSUMÉ EN FRANÇAIS

L'analyse d'incidents et d'expériences en laboratoire a montré que les états attentionnels dégradés comme la tunnélisation attentionnelle et la désorientation attentionnelle provoquent des réactions inattendues des opérateurs (pilotes, conducteurs, contrôleurs aériens, superviseurs de drônes par exemple). Une piste intéressante pour répondre à ces problèmes s'appuie sur les systèmes adaptatifs qui pourraient assister l'opérateur en temps réel, en changeant le comportement du pilote automatique par exemple. Ce type de systèmes adaptatifs requiert l'état attentionnel de l'opérateur en entrée. Le but de cette thèse de doctorat est de proposer des moyens de détection des états attentionnels en temps réel pour alimenter de tels systèmes, mais aussi pour fournir de nouveaux outils aux experts des facteurs humains. Pour cela, des méthodes automatiques d'analyse et des métriques des états attentionnels dégradés doivent être proposées. La mesure de l'activité oculaire qui est une piste particulièrement prometteuse fait l'objet de cette thèse.

Dans un premier temps, la vision humaine est présentée selon le parcours de la lumière et son intégration dans le cerveau. Le nécessaire filtrage de l'infor-mation est alors mise en évidence, justifiant d'étudier les mécanismes cérébraux de sélection, en particulier l'attention sélective. Différents modèles de l'attention visuelle en fonctionnement nominal sont alors mis en lumière, en commençant par une rétrospective sur les modèles classiques pour s'achever par le détail de quatre modèles récents. Les altérations de l'attention observées en environnement contrôlé sont ensuite présentées (tâche de détection de cible en laboratoire par exemple). Le lien est alors fait entre ces modèles théoriques et l'accidentologie des domaines d'intérêt tout en mettant l'accent sur la synergie entre atten-tion et vision que les travaux récents ne cessent de mettre en évidence. En particulier, il faut noter que les réseaux cérébraux de l'attention visuelle et de la vision sont quasiment confondus. La mesure de l'activité oculaire est donc un moyen pertinent pour la mesure de l'attention, car elle en est la principale manifestation. Il s'agit donc de se pencher sur les moyens de mesure de l'activité oculaire, dont on retient que la méthode de référence actuelle est la vidéo-oculographie. Son principe est détaillé, ainsi que le calcul des métriques oculaires parmi lesquelles certaines permettent l'étude de la tunnélisation attentionnelle: le temps passé sur les différents instruments du cockpit par exemple.

S'appuyant sur les travaux présentés jusqu'alors, la première proposition de ce manuscrit consiste en l'automatisation du traitement en temps réel des métriques connues de la tunnélisation attentionnelle pour sa détection. Ce travail s'appuie sur les données oculaires d'une expérience robotique ayant déjà été conduite au laboratoire. Cette expérience particulièrement pertinente pour notre propos, a été spécifiquement conçue pour favoriser la tunnélisation attentionnelle. Il est tout d'abord vérifié que sur les 23 participants de l'étude, deux groupes sont bien générés, un groupe dit "TUN=1" de 10 participants ayant fait face à la tunnélisation attentionnellle et un groupe contrôle "TUN=0" de 8 participants. Trois métriques connues de la tunnélisation attentionnelle sont donc calculées pour chaque participant, puis il est vérifié que ces métriques sont bien des marqueurs de la tunnélisation attentionnelle au travers d'une étude statistique comparative. Nous disposons ainsi d'une base de données qui associe les valeurs des trois métriques à un niveau binaire de tunnélisation attentionnelle. Il est alors possible d'utiliser une méthode d'apprentissage pour entraîner un classifieur à la détection des états de tunnélisation à partir des métriques calculées. La méthode retenue est un classifieur neuro-flou, qui obtient d'excellentes performances de classification et apporte une grande valeur ajoutée pour l'analyse des liens entre métriques et tunnélisation du point de vue des facteurs humains. Ce classifieur peut donc être utilisé pour automatiser la détection de la tunnélisation attentionnelle en temps réel en étant alimenté par les métriques pertinentes.

Néanmoins, il faut noter que les métriques utilisées pour cette classification sont spécifiques du contexte d'utilisation (il faut connaître l'environnement pour pouvoir faire le calcul des métriques). La stratégie est donc de ne pas utiliser ces métriques mais plutôt des métriques "directes", indépendantes du contexte. Mais il a été mis en évidence dans l'état de l'art que les algorithmes actuels de calcul de ces métriques présentent de nombreuses limites. Notamment, la cohérence des analyses est remise en cause dans la mesure où les algorithmes indépendants qui sont utilisés pour l'identification des événements oculaires se contredisent par instant. La deuxième contribution de ce manuscrit est donc un algorithme permettant le calcul unifié des métriques directes usuelles tout en garantissant la cohérence de l'identification. Son fonctionnement est équivalent à celui d'une machine à état, ce qui justifie son nom: Algorithme d'Identification des États Oculaires (ESIA dans le texte). L'approche est ensuite validée au travers de plusieurs comparaisons et analyses de performances sur plusieurs jeux de données provenant de contexte différents: supervision de robot à distance, pilotage en simulateur simple ou complet Airbus/Boeing avec équipages professionnels ou encore tests psychotechniques des Élèves Pilotes de l'École Nationale de l'Aviation Civile. Une nouvelle métrique est alors proposée en s'inspirant des

travaux précédent, le ratio exploration versus exploitation de l'information visuelle. Des deux cas d'étude mette en pratique les métriques calculées grâce à l'ESIA pour l'estimation des états attentionnels dégradés. L'expérience robotique, dont les données ont été utilisées pour entraîner le classifieur neuro-flou à détecter la tunnélisation attentionnelle constitue le premier cas. Il est alors montré que le ratio permet d'identifier statistiquement les participants du groupe TUN=1 et du groupe TUN=0, et qu'il est donc une métrique pertinente de la tunnélisation attentionnelle. Puis cette approche est reproduite en simulateur de vol dans le cadre d'une étude du Bureau Enquête Analyse (BEA) mettant en évidence l'intérêt du ratio pour analyser rapidment des enregistrements de plusieurs heures. Le deuxième cas d'étude concerne la détection de la désorientation attentionnelle: une expérience de pilotage au pilote automatique en simulateur de vol permet de provoquer la désorientation attentionnelle de plusieurs pilotes équipés d'un oculomètre. Les métriques oculaires de ces pilotes sont calculées et comparées sur deux périodes, une période de référence, et une période de désorientation attentionnelle. Il est ainsi démontré statistiquement que le ratio est également un marqueur de la désorientation attentionnelle.

Le travail est finalement discuté, notamment son intérêt en regard de la problématique de facteurs humains présentée en introduction. On pourra retenir que l'automatisation des états attentionnels en temps réel est possible et que les méthodes neuro-floues sont particulièrement adaptées: elles sont robustes aux entrées psycho-physiologiques bruitées et continues, et permettent de traduire leurs lois d'inférence en langage naturel. Ceci permet de les comparer aux lois utilisées par les experts, pour la détection de la tunnélisation attentionnelle notamment. Par ailleurs, en affinant l'analyse de l'activité oculaire au travers de l'ESIA, il est possible de calculer en temps réel des métriques de l'état atten-tionnel qui ne dépendent pas du contexte d'utilisation. L'ESIA permet un gain de temps triple, aucune paramétrisation de l'environnement n'est nécessaire, toutes les métriques usuelles sont calculées en une seule exécution de l'algorithme en temps réel et les métriques calculées permettent d'isoler rapidement quelques secondes critiques au sein d'enregistrements de de plusieurs heures. De nombreuses pistes d'amélioration sont finalement proposées, tant du point de vue du classifieur que de l'ESIA. Il est également souligné qu'un développement approfondi doublé de nombreux tests et certifications seront nécessaires pour porter ces méthodes sur le terrain, tout particulièrement dans les domaines d'intérêts tels que l'aéronautique, l'automobile, le contrôle aérien ou la supervision de drônes.

ENGLISH VERSION

It is largely accepted that pilot errors represent a major cause of aircraft crashes, being more frequently cited than mechanical failures (Inglis et al., 2007, Lourens, 1989, O'Hare et al., 1994). Safety statistics show that the gradual introduction of automation and assistance since the 1960's has improved safety, with modern « computerized» cockpits taking pride of place in an accident rate which is half of the rate of the previous generations of aircraft (Boeing, 2012). However it appears that such technologies and high operational pressure have created a new category of potentially deadly incidents in which crews are unable to comprehend the situation that they face and persist in erroneous decision-making. The development of automation has dramatically changed the role of the human operator from "direct (manual) controllers" to a "system supervisor / decision maker" (Parasuraman and Wickens, 2008): nowadays, pilots handle the stick for less than five minutes per flight on average. Increased trust in automation reduces the crew's basic flying abilities, and leaves them ill equipped to cope with emergency when automation fails (Mumaw et al., 2001). The complexity of the modern flight deck and the absence of appropriate feedback from human machine interfaces (HMI) can overwhelm even the most experienced human operators when the unexpected happens. Occasionally pilots are confused and persist in erroneous courses of actions that conflict with flight management despite critical visual and auditory alerts. It is expected that the occurrence of such accidents will increase in the future when "Single-Pilot Cruise Concept" will be introduced for long-haul operations. The concept holds that a reduced crew would be required to remain in control for extended periods during the cruise phase while the other crew members would be resting outside of the flight deck. If a major failure requiring an emergency descent occurs, a single pilot would have to operate the aircraft, communicate with air traffic controllers, manage the failure and perform the procedure, all whilst wearing an oxygen mask. These considerations are motivated by the recent expansion of unmanned air vehicles for military and civilian purposes, but human factors concerns raised by such paradigms are of great extent (NASA-ARC, 2012). One should consider that the accident rate of these highly automated systems are very high: for example, the Customs and Border Protection agency (BPC), which flies unmanned systems on border patrols, has an accident rate that is more than 353 times the commercial aviation accident rate (1 accident every 1850 flight hours) despite the use of a safety ground pilot (Haddal and Gertler, 2010). Worst, another report from the MIT International Center for Air Transportation points out an accident rate 5000 times higher for UAVs than commercial aviation (Weibel and Hansman, 2006). One must admit that commercial aviation is mainly

motivated by economic considerations, potentially to the detriment safety, and will put more and more pressure on the human operators who will have to adapt to these systems. Human Factors will therefore remain of major importance in order to maintain safety in the context of human-system interactions.

Traditionally Human factors studies have focused on descriptive theories and promoted behavioural approaches to assess the pilot's performance. Complementary subjective measurement techniques such as questionnaires have been proposed to get direct feedback from the operators. For instance, the NASA-TLX, SAGAT, SART are standardized workload and situation awareness questionnaires. Though they allow operational conditions or interface layouts to be compared, they have also strong limitations. Indeed the use of questionnaires is not convenient as it requires either to interrupt the task to get immediate feedback or to ask participants after the end of the experiment. Furthermore, questionnaires reflect the subjective analysis of the operators, which is impacted by interpretations and memory biases. Such methodologies have provided valuable insights into the operator's activity, but they are somehow limited to understand the underlying cognitive mechanisms of human error or to assess the cognitive load in highly dynamical contexts. As far as the AF447 Rio-Paris is concerned, it is still hard to understand why the pilots did not react as expected to the stall alarm and persisted in erroneous schemes of actions. Not only the interaction between the human and the machine is at stake, but also the human inner functioning. The operational conditions might indeed push operators to the limit of their perceptual, cognitive and executive capabilities. Better understanding and modelling these limits and their consequences is a major concern. Indeed, many authors have directly linked these impairments with the non detection of unexpected events, or potential hazard in turn responsible for the deadliest accidents in aeronautics (Shappell and Wiegman, 2003, Inglis et al., 2007), automotive (Velichkovsky et al., 2002, 2005, Martens, 2011) or in unmanned vehicles supervision (Waraich et al., 2013). An innovative and original approach coming from neurosciences and psychophysiology is to use objective measurements as input of an automated inference system to adapt the system and its automation in real time in order to cope with the operators impairments during critical situations as proposed by citeParasuraman2008.

Measurement techniques such as Eye-Tracking have indeed proven their ability to investigate the visual attention mechanisms since the pioneer work of Yarbus et al. (1967). Eye-tracking is now an affordable well-mastered technology. It is portable in the field via head-mounted systems, and allows for very precise and fast sampling of the tiniest movements of the eye. Some devices are not intrusive and do not interfere with the task at hand. Many research laboratories are now equipped with eye-trackers, but also neuro-marketing companies, hu-

man factor consultants, or even eye surgery blocks. Though the technique is mastered and basic metrics have been proposed Holmqvist et al. (2011), it is worth noting that the analysis of the data is not standardized and its use by human factors experts is limited (e.g. "heatmaps"). Most of the eye-tracker holders leave it up to the manufacturer for the basic analysis of the ocular activity, which in turn has a dramatic impact on the quality of the overall results based on eye-tracking. As a consequence, the subsequent analyses might not be reliable. Handling the data produced by an eye-tracker remains a challenge, which opens numerous opportunities for improvement and innovative solutions to be proposed.

The purpose of this Ph.D. thesis is to provide tools and metrics based on oculometry to measure the operator attentional state. Our goal is to support humans factors experts in designing systems which would be more robust to human attentional impairments in critical situation. An emphasis in put on making these tools and metrics not only automatic and usable in real-time, thus paving the way to adaptive systems and automation, but also context independent in order to be used seamlessly not only in planes but also in cars, supervision interfaces for drones, or even video games.

The Chapter 2 of the dissertation presents the *human vision*. All steps of the process are detailed following the integration of light in the brain: the light entering the eye, its conversion in the retina, the visual feature identification in the visual cortex, the separation between the dorsal and ventral pathways, and finally the orienting of the eye by the ocular plant.

The Chapter 3 addresses the issue of selecting the relevant information out of the environment performed by the *visual attention*. It starts with a retrospective on the historical models of attention and goes to the most recent models. The emphasis is put on four models. Posner's model of the orienting of attention covers the brain regions that are involved in shifting the focus of attention from one target to another. Michael's Master activation Maps consider the problem of spatial selection of the relevant stimuli, whereas Wicken's SEEV proposes a probabilistic approach to this issue. The last model from Findlay addresses the temporal aspect of the attentional selection. These models give insights for understanding the attentional impairments, which are presented in the last section. The focus is put on the reduction of the functional field of view, inattentional and change blindness and attentional tunnelling, all of which threatening the success of the operators' mission.

Eye-tracking techniques are covered in the Chapter 4. Is the apparatus detailed and the ocular data flow that goes through computation, analysis and results of the ocular metrics from which attentional

markers are derived. Based on this review of the current knowledge and literature, we propose the following contributions.

The Chapter 5 consists of the first contribution of this Ph.D. From data that were previously recorded during a robotic experiment, we have implemented an *automated attentional tunneling inference system* that can identify the attentional tunnelling periods of an operator from the analysis of their ocular activity and cardiac rhythm. This experiment was designed for provoking attentional tunnelling: two groups were isolated, one that faced attentional tunnelling for periods that we identified, and one that did not. In parallel to the mission, the cardiac rhythm and ocular activity were recorded. A machine learning technique named ANFIS elicited the links between the metrics and the attentional state. This classifier was then tested on new data and proved that the approach was relevant.

The ocular metrics that we used in the first approach required the interface to be known. Therefore we have investigated another kind of *ocular metrics that do not depend on the interface* in Chapter 6. In order to compute such interface-independent metrics, we have proposed an algorithm (ESIA) that overcomes some limitations of classical ocular event identification. The algorithm was first validated, and then tested on three use cases.

In Chapter 7, with the data from the robotic experiment, we have demonstrated that the exploitation / exploration ratio computed from the ESIA allows the participants who faced *attentional tunnelling* to be identified. The following French Safety Board (BEA) study has demonstrated that the ESIA ratio could identify periods of suspect attentional states that were further confirmed as dangerous situations by their human factor experts.

Finally, in Chapter 8, the exploration / exploitation ratio was proven efficient in detecting the pilots' *attentional disorientation* when facing "automation surprise".

The work is finally discussed in Chapter 9.

2

THE HUMAN VISUAL SYSTEM

RÉSUMÉ EN FRANÇAIS

Ce chapitre introduit les mécanismes principaux de la vision humaine. La vision humaine est un système extrêmement complexe de part sa capacité à percevoir les couleurs, les formes, les mouvements, les distances, et à intégrer le tout en une représentation unifiée de l'environnement. Pour simplifier, ce système est composé d'un senseur de lumière, d'un système de traitement de l'information, et d'un système d'orientation. Les parties de ce chapitre s'enchaînent tel le chemin d'intégration et d'interprétation de la lumière dans le système nerveux central, pour finir par les structures permettant l'orientation des yeux.

L'oeil est l'organe qui permet de percevoir la lumière provenant des objets de notre environnement. Dans un premier temps l'oeil joue un rôle de conditionnement de la lumière. La structure et les mécanismes d'adaptation de l'oeil sont faits de sorte qu'une image « nette » de l'environnement se forme sur la rétine. L'intensité lumineuse de cette image est contrôlée par la pupille pour l'adapter aux seuils de fonctionnement des photo-récepteurs rétiniens.

Les bâtonnets et les cônes sont deux types de photo-récepteurs rétiniens dont le rôle est de convertir la lumière les excitant en impulsions électriques. Dès lors, l'information lumineuse ne transitera plus que sous forme d'impulsions électriques dans le système nerveux central. Un premier traitement de cette information a lieu dans la rétine elle-même, dans la couche dite d'interconnexion. Les neurones de cette couche font converger les données de plusieurs photo-récepteurs. Leurs répartitions hétérogènes dans la rétine permettent d'expliquer la dualité de performance du système visuel humain: haute résolution au centre du champ de vision, bonne sensibilité aux mouvements en périphérie. Le traitement de l'information visuelle continue dans la couche ganglionnaire qui assure notamment la détection de contour. On note deux types de ganglions, M et P, qui projettent sur deux canaux nerveux distincts qui constituent le nerf optique, respectivement, M et P.

L'information visuelle sort donc de l'oeil par ces deux canaux qui projettent respectivement sur les couches M et P du Noyau Geniculé Latéral (NGL). La couche M (correspondant à la périphérie du champ de vision), insensible à la couleur, réagit rapidement aux stimulations lumineuses de grande taille et est sensible aux mouvements, alors que la couche P (correspondant au centre du champ de vision), sen-

sible à la couleur, réagit lentement aux stimulations lumineuses de petites tailles, et n'est pas sensible aux mouvements. Ces caractéristiques sont maintenues après projection du NGL vers le cortex visuel primaire V1.

Les caractéristiques telles que la fréquence spatiale ou les orientations sont calculées dans V1. Puis V2 permet d'enrichir encore l'analyse de la scène visuelle en combinant les caractéristiques spatiales de plusieurs régions: la "rétinotopie" est alors perdue. Les informations visuelles vont ensuite emprunter deux chemins distincts, ventral et dorsal.

Le chemin ventral emprunte les cortex visuels V3 et V4 dont le rôle est essentiellement la reconnaissance d'objet pour une interprétation sémantique (le « quoi »). Dans le modèle focal versus ambiant, le chemin ventral correspond au mode focal de la vision: les fixations longues permettant l'encodage minutieux d'une zone restreinte de l'espace sont combinées avec des petits déplacements oculaires (saccades), permettant de couvrir tout l'objet examiné. A l'opposé, le chemin dorsal projette directement sur MT/V5 où les mouvements et dispositions des éléments de la scène sont encodés, plutôt que leurs significations. Ce chemin correspond au mode ambiant: une succession de mouvements de grande amplitude de l'oeil combinés à des fixations courtes permettent l'encodage des contrastes et informations globales d'organisation d'une grande partie la scène (le « où »). Autrement dit, le mode focal permet l'encodage des sources d'informations jugées pertinentes, alors que le mode ambiant permet la recherche de nouvelles sources informations.

Pour assurer ces deux odes de fonctionnement, le système visuel est doté du réseau musculaire oculomoteur. Celui-ci permet deux types de mouvement, les fixations et poursuites lisses, qui maintiennent de l'oeil sur une région d'intérêt, qu'elle soit respectivement fixe ou mobile. Il permet également le changement de région d'intérêt grâce aux saccades, mouvements rapides de l'oeil (pendant lequel l'oeil est aveugle).

Ce chapitre s'achève sur la mise en évidence des capacités limitées d'encodage de l'oeil, et sur la nécessaire sélection des sources pertinentes dans l'environnement pour optimiser l'exploitation de la vision. L'étude du processus de sélection des informations pertinentes est l'objet du chapitre d'après: l'attention visuelle chez l'homme.

ENGLISH VERSION

The goal of this chapter is to provide the reader with the essentials of how visual information is gathered and integrated. The human vision is a very complex system, for its ability to distinguish colours, shapes, motion, distance, and build rich representations of the environment. From a simplistic functional point of view, the HVS is composed of a light sensor, a signal processing system and an orienting system. The interested reader will enjoy reading (Filipe and Alexandre, 2013) as an efficient summary on humain vision, (Duchowski, 2007) as a reference book, and (Hubel, 1995, 2010) for a very comprehensive review of the neural correlates of the Human Visual System (HVS) and their implications in the visual information integration.

2.1 PRELIMINARY DEFINITIONS

VISION is the ability to perceive and analyse light from the outside world. Human vision covers the segment of the light spectrum ranging from 400 nm to 800 nm, called the visible spectrum of light.

EYE Organ of vision, working as a light conditioner and a light sensor. In the human eye, an inverted image of the outside world is projected on the retina. The retina, which is an extension of the central nervous system covering the inner layer of the eye, converts light into electrical impulses and extracts the basic features of the image that are carried out of the eye through the optic nerve.

ACTIVATION of a neuron denotes the emission (or firing) of stereotyped electrical pulses called *action-potential*. Activation can take the shape of one isolated pulse, or a chain of pulses at a certain firing rate. It is performed when the activation of the neuron inputs complies with the neuron specific function.

OPTIC NERVE The optic nerve is a myelinated (i. e. insulated) nervous channel that conveys the electrical signals gathered in the retina to the Lateral Geniculate Nucleus (LGN) through the magnocellular (M-) and parvocellular (P-) channels and to the Superior Colliculus (SC) through the SC-channel.

VISUAL CORTEX is the main visual processing system that builds a representation of the outside world provided to the other structures of the brain through the ventral and dorsal streams. The dorsal stream innervates the Post Parietal Complex (PPC) that is thought to be involved in fixations.

OCULOMOTOR SYSTEM denotes the structures of the brain and the muscles around the eye programming and performing the eye

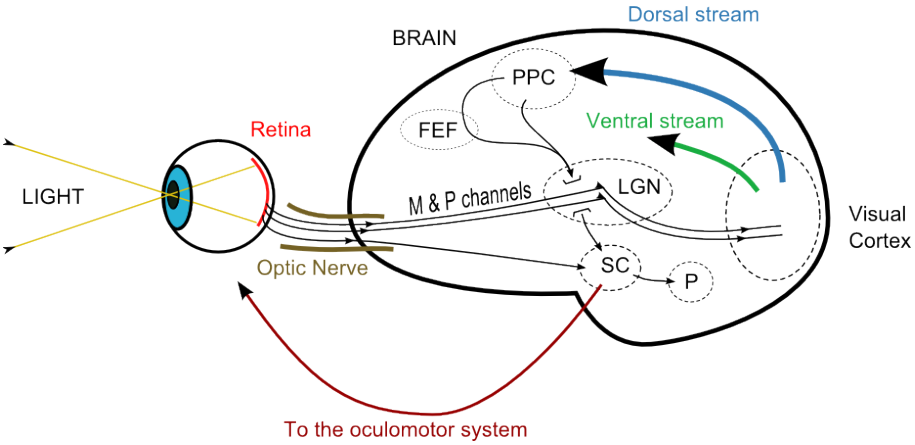


Figure 1: Simplified representation of the human visual system. Please refer to Section 2.1 for acronyms definitions.

motions for orienting the eye balls. The SC is involved in programming eye movements, but is also connected to the pulvinar (P) which helps in focusing attention.

Vision in ensured thanks to the eye that conditions and converts the entering light to electrical signals that are carried in the brain through the optic nerve in the LGN and the SC. The LGN then projects on the visual cortex whose processed visual representation is used by the oculomotor system to orient the eye. The structures and paths of the human visual system are depicted in Figure 1.

2.2 LIGHT CONDITIONING : THE EYE

This section gives a simplified presentation of the eye by restricting the content to information that is useful in explaining the functional features of human vision. A representation of the human eye model considered in this section is presented in Figure 2.

The white and opaque outer surface of the eye is called the sclera. The small region of the outer surface of the eye that is transparent and allows light to transit to the iris is called the cornea. The iris is responsible for controlling the amount of light entering the eye through the lens. The lens is actuated by the ciliary body so as to form a sharp image on the retina, which is the light sensitive region of the eye (Gregory, 1997).

2.3 EARLY STAGE OF VISION: THE RETINA

The eye conditions the entering light like the focus and aperture of a camera. The retina can be thought of as a frame of pixels but is misleading (Duchowski, 2007).

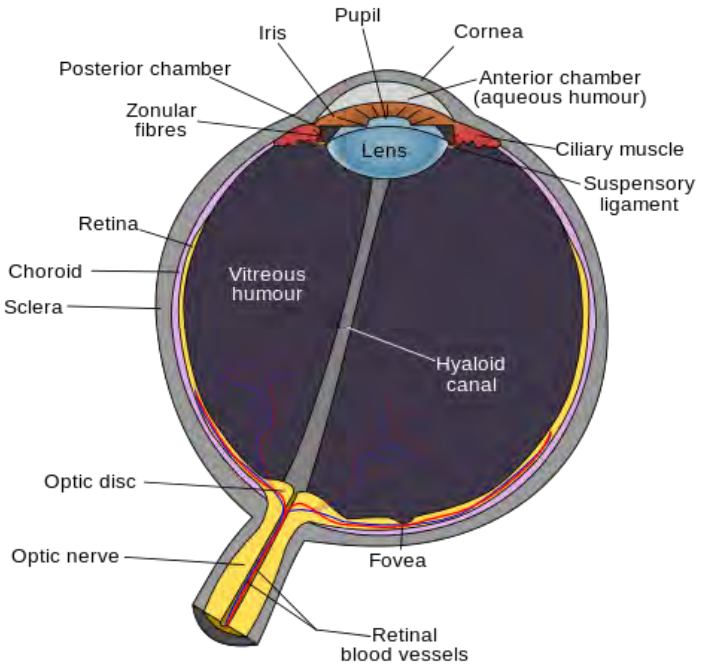


Figure 2: Human eye model. From Wikimedia Commons

First, the retina "pixels" are not sensitive to the same features all over the eye, and their layout is not uniform in the retina. For instance, the fovea and parafovea are very different in composition and function from their periphery. These small concentric areas of the retina are exposed to a very reduced portion of space: a 2° cone for the fovea, and 5° for the parafovea, whereas the periphery covers more than 120° of the visual scene.

Second, the retina "pixels" are interconnected to perform an early image feature detection, as edges for instance. The retina is in fact composed of three layers of neurons (see Figure 3) whose functions are detailed in the next sections (note that in the eye, neurons are called cells).

2.3.1 *Light conversion: the outer nuclear layer*

The outer nuclear layer is composed of rods and cones that are the light sensitive neurons of the eye. One surprising fact when considering the structure of the retina, is that the light receptors are at the outermost layer in the retina. Indeed, before reaching them, the light has to go through the two cells layers situated "above" (see Figure 3).

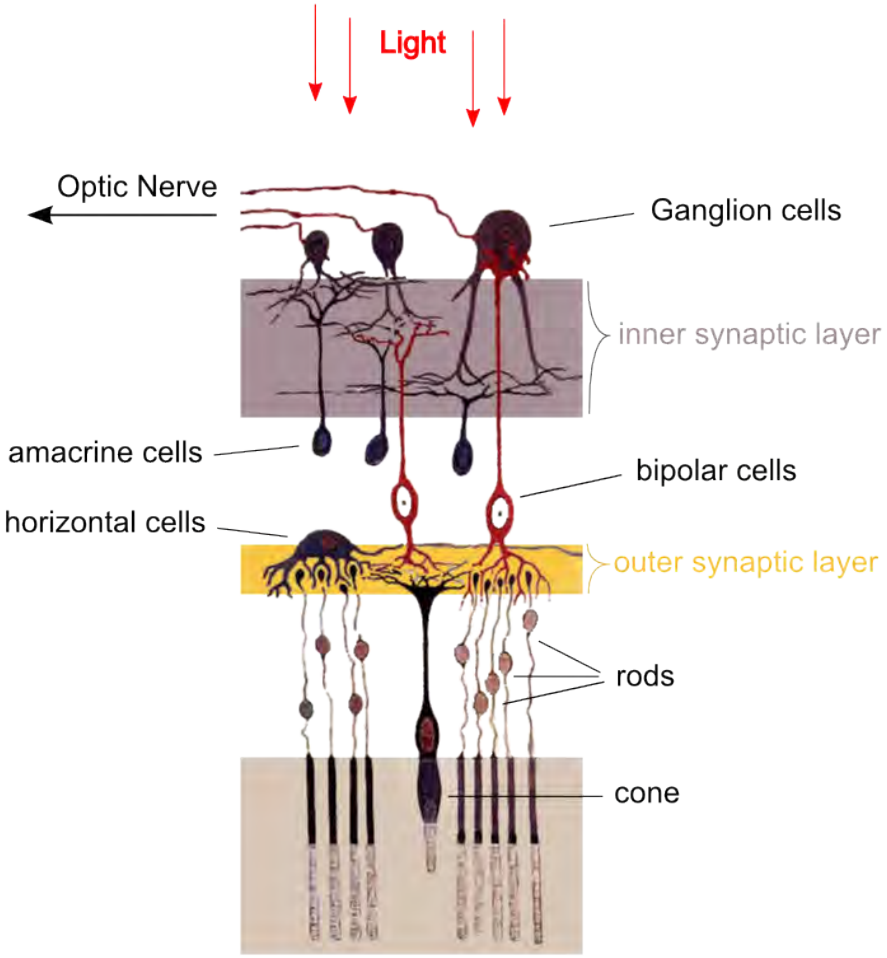


Figure 3: Human eye retina. Adapted from (Cajal and Azoulay, 1911).

Rods and cones can be seen as "transducers" neurons, converting the incoming light into an electrical signal (Cohen, 1972). To make it simple :

RODS are more sensitive to the range of light present at night allowing for monochromatic scotopic vision (i. e. under low light conditions) .

CONES are more sensitive to the range of light present during the day allowing for trichromatic photopic vision (i. e. in normal day light). They respond preferably to red, green or blue.

The light information is turned in electrical pulses inside the outer nuclear layer. The outer *synaptic* layer is the connection interface allowing the signal to transit to the inner nuclear layer.

2.3.2 *Interconnection: the inner nuclear layer*

The inner nuclear layer collects the signal from the outer nuclear layer via the outer synaptic layer. It is composed of amacrine cells, horizontal cells, and bipolar cells.

Bipolar cells transmit the electrical signals from *many* rods and cones to one ganglion cell in the fovea and several cells in the periphery. There is therefore a convergence of the signal in the eye (100 rods and cones for one ganglion on average), that can be seen as a data compression for transmission in the optic nerve (Filipe and Alexandre, 2013).

Horizontal cells have a rather slow response compared to the ganglion cells. They still contain the information of the previous position of an object when the new object position is detected and therefore play a key role in motion detection (Filipe and Alexandre, 2013)

In the fovea (see Figure 2), bipolar cells are connected to smaller group of rods and cones resulting in the central high resolution of vision. Out of the fovea, bipolar cells are connected to a wider number of rods and cones, resulting in the lower resolution and motion sensitivity of the peripheral vision (Duchowski, 2007).

2.3.3 *Receptive fields: the ganglion layer*

The interconnection of rods and cones in the inner nuclear layer results in what is called *receptive fields* in the ganglion layer. The ganglion fires only when the incoming light stimulates specifically the whole group of rods and cones to which it is connected. The area and characteristics of the stimuli necessary to trigger a ganglion firing is called the ganglion receptive field. One main characteristic of ganglionic receptive fields is their *center-surround* nature. It computationaly corresponds to contour detection: indeed uniform contrast

does not provoke ganglionic activation, as would a highly contrasted object (Hubel, 2010).

As presented in the previous section, bipolar cells in the fovea are connected to smaller group of rods and cones resulting in the central high resolution of vision. In other words, ganglionic receptive fields are smaller in the fovea than in the periphery. In general, receptive fields differ in size and characteristics all over the eye which results in ganglions with different functions (Croner and Kaplan, 1995).

Functionally speaking, there are three classes of ganglion cells: the X, Y and W, whose roles are to transmit the signals collected in the eye to the brain through the optic nerve in two different channels called the magnocellular (M-) and parvocellular (P-) respectively. (De Valois and De Valois, 1988)

X GANGLION CELLS respond to sustained stimuli, precise locations and fine details. They both project on the M- and P-layers of the Lateral Geniculate Nuclei (LGN).

Y GANGLION CELLS respond to transient stimuli, coarse features and motion. They project on the M-layer of the LGN.

W GANGLION CELLS also respond to coarse features and motion but project directly to the Superior Colliculus (SC).

2.3.4 *Conclusion*

The eye is by itself a complex system. It is able to condition the incoming light to form focused images of the outside world on the retina with controlled luminosity. The retina converts light into electrical signals, and allows for different features detection depending on the neurons interconnection. For instance, contrast or motion features are detected accordig to the X, Y or W nature of the ganglion. The visual signals go out of the eye through the optical nerve in 3 different M-, P- and SC-channels ensuring the continuity for the different features detection in the brain. We will now focus on the visual information transfer and integration in the brain.

2.4 VISUAL INFORMATION TRANSFER : THE VISUAL PATHWAYS

Ganglion cells transmit the electrical signals to the brain through the optic nerve. The optic nerve is the grouping of the ganglion cells long axon which is myelinated so that the electrical diffusion is faster from the ganglion to the brain. Both the right and left optic nerves converge at the optic chiasm, and that this crossing results in having the left visual field projected to the right hemisphere, and oppositely for the right visual field (see Figure 4).

As explained earlier, the optic nerve projection in the LGN is separated in the M and P channels. Linked with the differences in re-

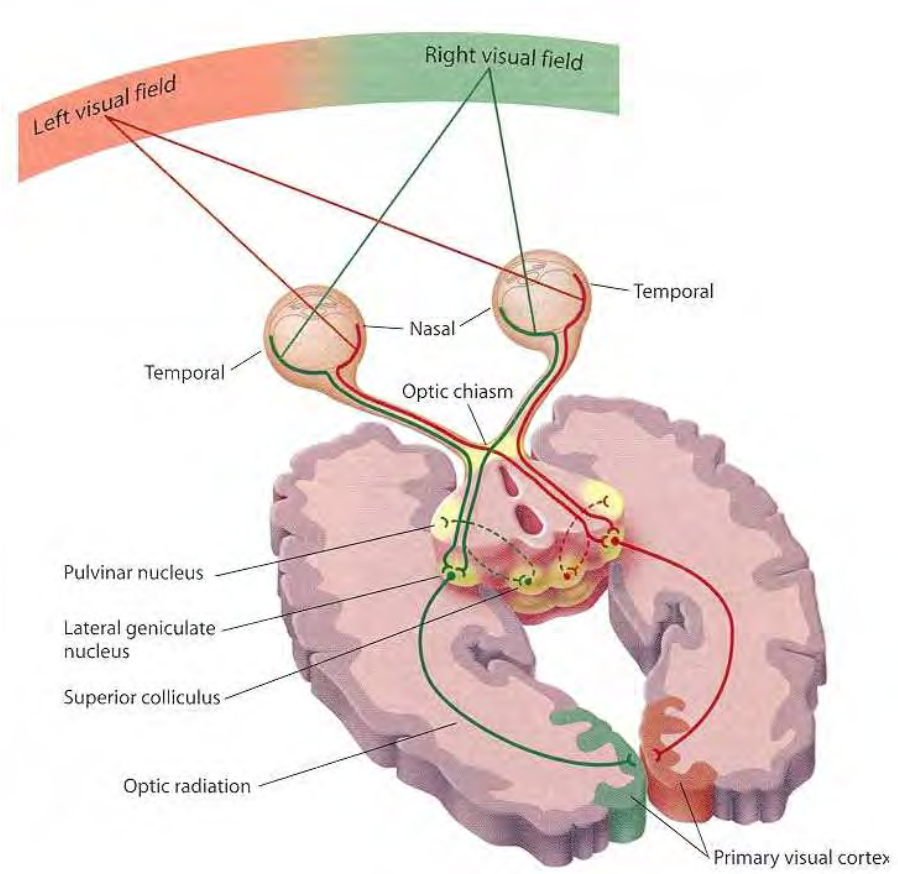


Figure 4: The visual pathways. From (Medical Atlas, 2012)

Characteristics	M-layer	P-layer
Main ganglion type	α	β
Ganglion size	Large	Small
Receptive fields	Large	Small
Sensitivity to small objects	Poor	Good
Sensitivity to contrast	Poor	Good
Transmission time	Fast	Slow
Sensitivity to light level changes	Good	Poor
Sensitivity to motion	Good	Poor
Color discrimination	No	Yes

Table 1: Functional characteristics of ganglionic projections. Adapted from (Duchowski, 2007).

ceptive fields in the ganglion layer, it is natural to observe that each channel is associated with specific functional characteristics of the ganglionic projections in the LGN. The functional characteristics of M- and P-ganglionic projections are presented in Table 1. As shown in Figure 4, M- and P-layers of the LGN are then continued in the primary visual cortex (V1).

2.5 VISUAL INFORMATION PROCESSING: THE VISUAL CORTEX

2.5.1 *Primary visual cortex: V1*

As explained earlier, the ganglion cells only fire when the light characteristics are compliant with the ganglion receptive field. Interestingly enough, the characteristics of the M- and P-channels is kept along the visual pathways to V1, with receptive fields in V1 being of the same resolution as the receptive fields covered by the ganglion cells. In other words, small receptive fields in the fovea are projected on small receptive fields of V1 allowing for a higher resolution analysis in V1. As ganglion cells are thought to be responsible for contour detection, V1 is thought responsible for the computation of *local* spatial features such as spatial frequencies, orientations, movements, directions (Filipe and Alexandre, 2013). V1 then project on the other areas of the visual cortex V2, V3, V4, and MT/V5.

In V2, the convergence of the receptive fields of V1 allows for a more complex representation of space, using spatial frequencies of different sub-regions for instance. The information is then separated in two streams that are located in the ventral and dorsal regions of the brain (Goodale and Milner, 1992) as presented in Figure 5.

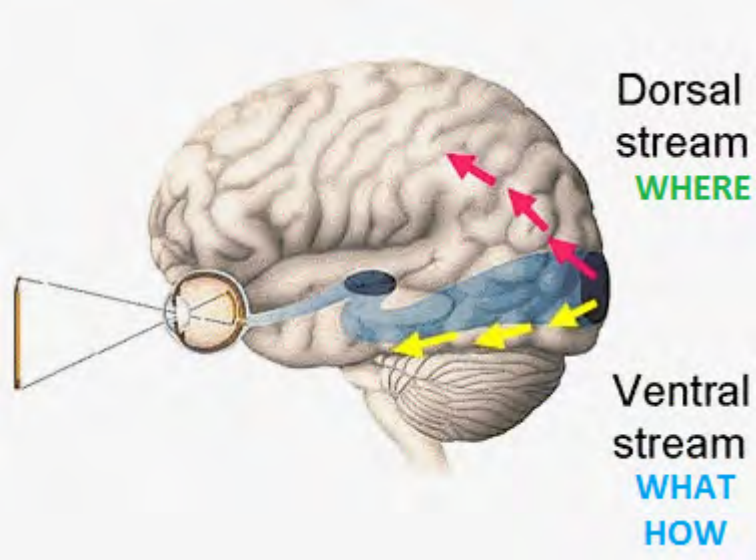


Figure 5: The ventral and dorsal pathways, associated with the attentional "what" and "where" (Press, 2011).

2.5.2 *The ventral (temporal) pathway: the "what"*

The feature analysis is furthered in V3 and V4 in the ventral stream and results in simple forms and colour detection (Zeki, 1993, Duchowski, 2007). Though controversial, this stream is now accepted to be providing attention with content related information (i. e. the "what"), using the form and colour feature detection to form complex and precise object recognition allowing for a meaningful representation of the environment (Goodale and Milner, 1992). This model lead to the focal versus ambient visual system (Velichkovsky et al., 2005). According to this model, fixation length and saccade amplitude are correlated : the focal mode that is based on the temporal stream would involve long fixations (longer than 280 ms) with small saccade within the parafovea (less than 5°long). This mode would therefore support the "what" integration.

2.5.3 *The dorsal (parietal) pathway: the "where"*

The other pathway coming out of V2 is the dorsal pathway that projects on the Middle Temporal (MT/V5) and the Middle Superior Temporal (MST) who are thought to play a major role in the perception of motion (Zeki, 1993, Duchowski, 2007). It is now accepted that this pathway allows for the attentional "where" integration with spatial representation of the environment and mapping of the different components of a visual scene used for triggering actions for instance (Goodale and Milner, 1992). In the focal versus ambient model of Velichkovsky et al. (2005), the parietal stream supports the ambient

mode that involves a succession of short fixation raging from 50 ms to 260 ms and high extent saccade in the periphery of the parafovea (longer than 5°). This simplistic view of the "what/where" pathways are still debated but nevertheless considered as relevant (Filipe and Alexandre, 2013).

2.5.4 *Conclusion*

This section has given an insight of how light detection from rods and cones *converges* in other neurons defining their receptive field. The convergence of these receptive fields in other neurons in turn defines more evolved receptive fields like spatial frequency. Reiterating this process numerous times in parallel allows for all the visual features to be "computed" and provides the brain with visual information that will be used as inputs for making decisions or trigger actions, like moving the eye to another location.

2.6 ORIENTING THE EYE: THE OCULOMOTOR SYSTEM

The oculomotor system allows for orienting the eye appropriately on what is the focus of attention. In that sense, it can act in two ways: maintaining the eye on a region of interest, or shifting to the next target of interest.

2.6.1 *Orienting actuators: the ocular plant*

The ocular plant is the network of muscles around the eye allowing for the eye motion. The ocular plant can act in three ways for orienting the eye verticaly, horizontaly, and around its own axis (twist). The muscles responsible for these 3 types of motions are respectively the superior and inferior recti, the lateral and medial recti, the superior and inferior obliques (see Figure 6).

These muscles perform the orienting movements of the eye according to commands that can be of three types (Duchowski, 2007):

VOLUNTARY when emanating from the occipital cortex (using visual information from the visual cortex)

INVOLUNTARY when emanating from the SC (using visual data from the SC-pathway before they are projected in the LGN and then processed in the visual cortex)

REFLEXIVE when emanating from the semicircular canals that provide the head attitude and motion

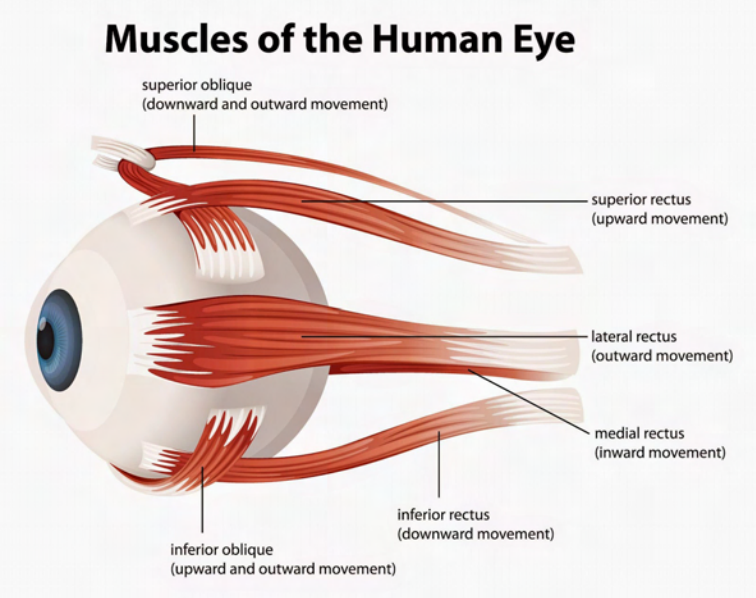


Figure 6: Muscles of the human eye. From (Pandey, 2013)

2.6.2 *Maintaining the focus on a region of interest*

Maintaining the eye on the target of interest requires a neuronal feedback loop circuitry. This loop pilots the eye for ensuring the stability of the target projection on the retina. Resulting motions are of various types:

FIXATION are relatively stable positions of the eye.

SMOOTH PURSUIT is a slow motion of the eye that compensates for target motion relatively to the head when slow enough (watching a cat wandering while you are seating, or a friend on the platform while your train is leaving the station)

Fixations are in fact characterized by very specific very low amplitude motions : *tremor*, *drifts* and *microsaccades*. We will see that these movements are naturally included in fixations when measuring the ocular activity at low frequencies. They are sometimes simply considered as noise (Duchowski, 2007) and will not be presented in this thesis (the interested reader can refer to (Ditchburn and Ginsborg, 1953)).

The feedback loop ensuring the stability of the eye on a target must be slow. It indeed requires processed visual data from the visual cortex, which is a long transit in the brain compared to the other modality of commanding the eye.

Fixations can be voluntary, also called top-down (TD), corresponding to a will to extract information from a specific region of interest allowing for processing in the visual cortex, or involuntary, also called

Type	Duration (ms)	Amplitude (°)	Velocity ($° s^{-1}$)
Fixation	200-300	-	-
Microsaccade	10-30	0.3 - 0.7	15-50
Tremor	-	0.03	0.3
Drift	200-1000	0.03-1	0.1 - 0.5
Smooth pursuit	-	-	10-30
Saccade	30-80	4-20	30-500
Glissade	10-40	0.5-2	20-140

Table 2: Typical values for the eye movements. Adapted from (Holmqvist et al., 2011)

bottom-up (BU) corresponding to a scan of the environment with reduced processing in the SC only.

2.6.3 *Shifting to another target*

Shifting the eye to another target is performed via *saccadic* movements. Saccades are said to be ballistic (you cannot correct a saccade once it is programmed) and stereotyped (profiles of saccades are known and repeatable).

The eye is assumed to be blind during the shift because it is to short to allow for visual processing.

Associated with saccades, *glissades* are oscillations of the eye around the desired final position (during the stabilisation phase).

Saccades can be voluntary or involuntary depending on the part of the brain taking control over the eye orientation. Again, voluntary saccades are based on processed visual information in the visual cortex (TD) and involuntary saccades use less processed visual information from the SC directly (BU). The Front Eye Field (FEF) is also thought to play a major role in eye orientation shifting.

2.6.4 *Eye movements characteristics*

Maintaining the eye in a stable position or shifting to another target therefore result in various types of movements that have specific characteristics and brain regions. Velocity, amplitude and duration of the eye movements have been described previously in (Holmqvist et al., 2011) and are presented in Table 2

2.7 CONCLUSION

The essential elements of the human vision system explaining the functioning and visual information integration of the main structures of the brain have been presented in this chapter. Vision is a tool providing sharp, parallel, distributed, and meaningful data to the brain used for our survival. But one has to consider the overly simplistic approach here, which hides the real complexity and still limited knowledge of the internal loops that modify the course of the visual stimuli.

Indeed, this presentation followed the *feedforward* direction of the visual information propagation as represented in Figure 5. That model does not account for the real complexity of the human visual system processing. Indeed neurons are not only connected to the "next" visual processing (V1 to V2, for instance), as some neural connections are directed in the opposite directions. These *feedback* connections impact the processing of the previous visual entity in the chain. In fact, the problem is even more complex, as neural connections are not restricted to these two directions, but are free to reach any other structures in the brain...

HVS is limited in capacity. Detailed information are provided only on the fovea and parafovea. It is therefore required to change the orientation of the eye sequentially to the other relevant objects in the environment. The choice of the next target remains a critical issue in vision. Our survival depends not only on the visual information processing but also on the selection of the relevant stimuli in the environment. This is the purpose of *visual attention*. It is a very debated research topic which is presented in Chapter 3

3

THE HUMAN VISUAL ATTENTION

RÉSUMÉ EN FRANÇAIS

Ce chapitre présente tout d'abord une rétrospective sur les modèles de l'attention visuelle. Il se focalise ensuite sur 4 modèles d'intérêt dans le cadre de cette thèse. Ces modèles servent ensuite de cadre pour expliquer les altérations du fonctionnement de l'attention qui sont observés dans de nombreux accidents en aéronautique notamment.

Le chapitre précédent concluait sur le fait que le cerveau a une capacité de traitement d'infor-mation visuelle limitée et sur la nécessaire sélection des stimuli pertinents pour optimiser l'utilisation des ressources cognitives. Dans les situations dangereuses, cette sélection devient cruciale pour permettre la prise de décision et l'exécution des actions adéquates dans le temps imparti. L'attention est classiquement catégorisé en 3 types : l'attention sélective (isoler une information des distracteurs environnants), l'attention divisée (suivre plusieurs objets en parallèle), l'attention soutenue (maintenir une réponse stable à un stimuli stable) en lien avec la vigilance. Nous nous intéressons ici à l'attention visuelle sélective seulement. L'attention et la vision sont classiquement différenciés par le fait qu'un changement d'objet de l'attention peut s'accompagner d'un mouvement des yeux (overt shift) ou non (covert). Par ailleurs, il est observé que ce changement d'objet peut être le résultat de stimulations extérieures (involontaires ou bottom-up) ou d'un contrôle conscient (volontaires ou top-down). Ces notions ont été formalisées au cours du temps, pour expliquer les différences entre les modèles proposés dans la littérature.

Par ailleurs se pose la question des critères utilisés pour la sélection d'une source d'infor-mation plutôt qu'une autre. Le chapitre précédent a permis d'identifier que le système visuel génère des informations sur le contenu (chemin ventral) et sur la localisation des objets (chemin dorsal). Ces réflexions convergent vers l'identification de trois composantes de la sélection attentionnelle : le "quoi": l'information est choisie pour son contenu, le "où": l'infor-mation est choisie pour sa localisation ou ses caractéristiques physiques, et le "comment": l'infor-mation est choisie en fonction son rapport à la tâche en cours. Plusieurs modèles de l'attention visuelle sont alors proposés. Le modèle de Posner (1989) présentent les structures cérébrales investies dans le changement d'orientation de l'attention. Notamment, il décompose le processus en trois étapes : désengagement de l'attention par le cortex postérieur pariétal, déplacement de l'attention par le col-

iculus supérieur, et ré-engagement de l'attention par le pulvinar. Il est intéressant de noter que ces structures sont communes au système visuel. La théorie d'intégration des attributs de Treisman (1980) explique comment à chaque lieu de l'espace est attribué l'ensemble de ses attributs aux travers de cartes indépendantes (couleur, orientation, contraste). Le modèle SEEV de Wickens (2007) est une approche statistique expliquant le choix des cibles pertinentes en fonction de leur Salience, des Efforts (ou coûts) pour les atteindre, de l'Espoir de gain qui leur sont associées, ainsi que la Valeur attribuées à ces cibles. Le modèle de Findlay (1999) explique finalement que le changement d'objet de l'attention résulte d'une compétition ininterrompue entre deux stratégies: rester sur le même objet, ou changer. Dès que l'option de changer devient la meilleure option, alors le changement est ordonné. L'attention s'oriente alors vers la meilleure cible au moment du changement, dont la pertinence a été évaluée parallèlement sur la base de ses attributs et de la tâche en cours. Autrement dit, le modèle de Posner explique comment le changement s'opère, Wickens explique vers quelle cible le changement s'opère, alors que Findlay explique quand le changement a lieu.

Ces modèles présentent le fonctionnement nominal de l'attention et de la vision, mais il est observé que dans de nombreuses situations, les pilotes ne réagissent de manière inattendue aux stimulations de leur environnement. Il est par exemple observé que la capacité à percevoir dépend de la charge visuelle ou cognitive ce qui sera identifié comme une réduction du champ visuel utile, (Williams,1985). Mais la perception n'est pas seule en cause dans l'omission de cibles, l'attention est mise en évidence par les modèles de cessité au changement (Simons and Rensink (2005)), et de cessité attentionelle (Mack and Rock 1998a) et la cessité attentionnelle soutenue (Simons,1999). De nombreux travaux viendront ensuite démontrer que l'attention peut-être trompée de sorte que des événements mêmes saillants, attendus et pertinents à la tâche peuvent être négligés : c'est la tunnélisation attentionnelle (Wickens,2005). Ces phénomènes sont particulièrement pertinents pour expliquer le comportement des pilotes dans des conditions de stress et de charge mentale importante rencontrées dans les situations dégradées (panne, mauvaise météo). La tunnélisation attentionnelle est la cause de presque tous les accidents mortels de type CFIT (collision avec le sol en vol piloté) de l'aviation générale sur la période 1990 à 1998 d'après Shappell (2003).

Le chapitre conclu sur l'intérêt d'étudier ces phénomènes attentionnels pour éviter certains types d'accidents mortels notamment dans l'aviation. Le concensus stipulant que bien que attention et vision soit distincts, les changements attentionnels sont très majoritairement accopagnés d'un mouvement des yeux est rappelé. L'oculométrie (étude du mouvement des yeux) est donc un instrument particulièrement pertinent pour l'étude de l'attention appliqué aux facteurs humains.

ENGLISH VERSION

The brain has a limited infor-mation processing ability. Nevertheless, in our everyday life, we are surrounded by visual stimuli of all sorts. The straightforward way for maximizing the outcome of our brain visual processing is to select relevant stimuli out of the flood of information.

Attention is demanded in our everyday life, but also in critical situations. When it comes to handling risky situations, the selection of relevant stimuli, decision making and execution of actions are of primary importance. Attention is supposed to support these interventions but has its own limitations as it will be highlighted in this chapter.

This chapter aims at presenting the models of attention that have been proposed in the literature. First, the common grounding is presented so as to introduces the following retrospective on models of attention. We will then focus on 4 models that are of particular interest for understanding the attentional impairments presented in the last section.

3.1 PRELIMINARY CONCEPTS

As James (1890) said "*Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others. When things are apprehended by the senses, the number of them that can be attended to at once is small.*" The concept of attention has been widely studied in cognitive psychology where it is clearly associated with an infor-mation filtering mechanism Broadbent (1958) that allows to focus, divide or sustain attention.

FOCUSED ATTENTION (FA) is the ability to focus on a specific set of stimuli to the exclusion of distractors;

DIVIDED ATTENTION (DA) is the ability to process stimuli and tasks in parallel;

SUSTAINED ATTENTION (SA) is the ability to maintainone's behavior continuously in response to a continuous set of stimuli. This concept is generally linked to vigilance.

Von Helmholtz and Southall (1925) were the very first to study attention at a scientific level. They were interested by the continuous "roaming" of the eye in the environment. One major observation they made was that it is possible to focus on peripheral infor-mation without having the eye to move. This major and very debated concept in the study of attention was later called *covert* shifts of attention. It means that the attention can be focused without requiring the eye to

be oriented in that direction. Nevertheless, the eye moves very frequently in new directions, as it was also observed by Von Helmholtz and Southall (1925). They concluded that the movements of the eye allow for detailed vision of specific regions of the environment, "because that is the only way we can see as distinctly as possible all the individual parts of the field in turn". In contrast, an attention shift that is accompanied by an eye shift is called *overt*. This is summarized in the following definitions :

OVERT SHIFT is a change of the focus of attention that is accompanied by a shift in the eye orientation. In other words, the attentional shift is *o*-pen.

COVERT SHIFT is a change of the focus of attention that is *not* accompanied by a shift in the eye position. In other words, the attentional shift is *cover*-ed.

The covert nature of attention is generally accepted: sport is a typical situation during which overt shifts occur. Some authors like Posner (1980) gave a predominant role to covert shift in the orienting of attention. On the contrary, a more recent model of special interest adopted a radically different position, according very little contribution to covert shifts in the orienting of attention Findlay and Walker (1999). Both models are presented in detail in the next sections.

Von Helmholtz and Southall (1925) not only highlighted covert shifts of attention but also proposed very valuable insights on the control of attention:

1. Attention wanders to new things spontaneously as shown by the eye movements.
2. Conscious control is possible for choosing to focus on a precise location.

Though this was later formalized with this terminology, this introduced the concepts of bottom-up and top-down controls that are at the core of the study of attention. Bottom-up and top-down controls are defined as follows:

BOTTOM-UP (BU) means that reactions are driven by the environment stimuli. "Bottom" here denotes the environment, and "Up" denotes the brain highest level of cognition. Synonyms : involuntary, automatic, reflexive, stimulus driven, peripheral or ambient (using the fovea periphery neurons), exogenous (*gen*-erated by the *ex*-terior). BU stimulation propagates through feedforward connections.

TOP-DOWN (TD) means the Top (related to high level goals and control) applies a regulation on the reactions to stimuli (Down).

Synonyms : voluntary, controlled, goal oriented, central or focal (using the fovea neurons), endogenous (*gen*-erated by the *in*-terior). TD control is performed via feedback connections in all brain regions.

3.2 HISTORICAL REVIEW OF ATTENTION

Attention was first studied as a branch of Philosophy. Psychology then considered attention for its behavioural implications and underlying cognitive processes. Nowadays, attention is mainly explored under the light of Neurosciences. From the analysis of brain structures and their functioning, new models of attention have been proposed recently, and some of them implemented computationally (computational neurosciences).

3.2.1 *Selective attention*

The most famous illustration of selective attention is the "cocktail party effect" that was proposed by Cherry (1953). He noticed that one can select a conversation among many and follow it, discarding the others. Using *dichotic* hearing paradigm, he then showed that gender, direction, pitch and rate of speech impacted the ability to separate the messages. He carried on his experiments with the *shadowing* task. Two messages were presented at the same time in separated channels using a binaural head-set. The participants had to repeat aloud the message from the attended ear. The participants were successful in this task showing that attention is a selective filter. Another trivial observation is that our actions follow only one scheme of actions at a time, despite the fact that our senses perceive infor-mation that might ask for contradictory actions. From that observation, Broadbent (1958) hypothesized the existence of a "bottle-neck" between perception and action and a "Selective Filtering". This model formalized the notion of limited capacity, but also the filtering approach for the selection of the relevant stimuli going through the attentional "bottleneck". It is a theoretical approach, based on a filter between the sensory inputs that are stored in a "buffer", and the short-term memory used for processing infor-mation. His experiment provided the subjects with simultaneous auditory stimuli "123" et "abc". They were then asked to recall the auditory messages. The subjects answered "123-abc" or "abc-123" but no "1a2b3c" for instance. He concluded that infor-mation enters in parallel in the sensory channel buffer but is then selectively filtered to be processed sequentially, thus preventing an infor-mation overload in the high-level processing network of the brain. The selection would therefore be based on "where" the infor-mation comes from rather than "what" it contains, because the

Figure 7: Simple example of target selection. Compared to (a), reaction times measured for selecting the purple triangle in (b) are longer and grows with the number of purple elements in (b). This advocates for early perceptual filtering on the color feature that is sufficient for selecting the relevant target in (a) , whereas the perceptual filtering isolates four shapes that are processed sequentially for selecting the triangle among the other purple shapes in (b). From (Zawidzki and Bechtel, 2002)

selection is done *before* processing, which is also known as the *early selection* hypothesis.

Early selection was demonstrated with simple experiments as the one presented in Figure 7. But this theory faced controversy quite rapidly. Back to the cocktail party effect, Cherry noticed that while people were focused on one ear, "filtering" the other, they would notice their name even though it was emitted in the filtered ear. Moray (1959) went on with this observation, using the same shadowing paradigm to explore what kinds of messages could go through the filtering. He found out that messages of "subjective importance" are likely to be interpreted whereas no other message could go through. The contents would be used for selection *after* processing. Deutsch and Deutsch (1963) also contradicted Broadbent's model, because the theoretical filter would be necessarily at least as complex as the processing network of the brain for identifying right and wrong stimuli. Instead they proposed a model based on weights of *importance* in a central structure. In their model, all sensory inputs are processed at a high level, but some of them are attenuated according to their weights of importance. In this model, it is the "what" that is evaluated for selecting the most important content. This advocated for the *late selection* hypothesis which consists in having selecting after that sensory messages are processed.

As both late and early selections were supported enough, Treisman (1960) proposed a mix of both theories. She integrated an "attenuation filter" corresponding to Broadbent's selective filter, allowing for isolation of sensory messages and "dictionary units" that are weights used to reject weakened stimuli after the attenuation filter depending on their importance. This model brings both "where" and "what" cri-

teria to be part of the selection process, and suggests that both early and late selection steps are mutually performed and updated.

The filtering of relevant stimuli is a critical aspect of attention. But another controversial topic has brought new models of attention: the scene integration.

3.2.2 *Scene integration*

The scene integration denotes the construct of a relevant mental representation of our environment. How is it built ? One interesting stream in psychology came from the Gestalt Psychologists. They hypothesized that the scene is perceived as a *whole* in a parallel *one step* process emphasizing that "The whole is other than the sum of the parts" i. e. the whole is not integrated from sub-parts of the environment gathered together. This very strong hypothesis was invalidated by studies using the first continuous measurements of the eye movements for the study of attention. Indeed Yarbus et al. (1967) used eye-tracking to account for the path followed by the eye when inspecting simple images. They defined the concept of "scanpaths" as the sequence of the successive eye fixations and observed that :

1. A sequence of eye fixations is required to integrate the global scene, which invalidated the Gestalt hypothesis.

2. Visual scan paths are not the same when asking people to analyse different aspects of the same image.

This advocated for a sequential viewing of a scene with step by step integration of the "what" infor-mation for forming a bigger picture. The scene integration process therefore seems to be sequential in terms of "what", but performed in parallel in terms of "where". How is this representation of the environment stored ? This question raised the concept of mind maps, which has been used in more recent models of attention, including models dealing with attentional selection. Mind maps are theoretical tools for understanding the attention functioning.

3.2.3 *The "what", "where", and "how" of attention*

As explained in Chapter 2, ventral and temporal pathways have functional differences. The ventral stream is involved in the in-depth analysis of the visual stimuli in order to extract content related information (the "what"), whereas the dorsal stream is specialized in dealing with spatial feature extraction as localization and velocity (the "where"). This fundamental separation is also used in the study of attention. To some extent, James (1890) highlighted the "what" of

attention saying that internal thoughts were the nature of attention, whereas Von Helmholtz and Southall (1925) focused on the "where", observing the attentional selection of stimuli in space, whether overtly or covertly.

Gibson (1941) added the fact that intentions pre-condition the outcome of attentional processing. He is especially known for his work on manipulation of expectations, which proved the "how to interpret" factor of attentional processing. For instance : the word "sael" quickly presented to a viewer in the context of sailing ends up being perceived as "sail"; whereas in the context of animals, it ends up being perceived as "seal". This notion is the precursor of the "attentional-set" of Theeuwes (1994) that is still current.

ATTENTIONAL "WHAT" denotes the content and meaning of the information provided by the attended element. In order to be encoded from the visual scene, it requires the processing of the visual cortex and further analysis in the ventral stream.

ATTENTIONAL "WHERE" denotes the location and motion of the attended element. It is processed in the dorsal stream.

ATTENTIONAL "HOW" denotes the intentions that will impact the future reactions to the infor-mation provided by the attended element.

These models of attention remain relevant as attention is still studied in terms of TD/BU covert/overt, "what", "where", and "how", even though these concepts have changed names in the course of history. We therefore propose some major models that implement the underlying mechanisms of attention.

3.3 MODELS OF ATTENTION

Recent research argues in favor of a distributed attentional selection in vision structures. Filipe and Alexandre (2013) put it straight:

> "The non existence of a brain area solely oriented for visual attention is one of the most important discoveries in neurophysiology, but the visual selection appears to be present in almost all areas of the brain associated with visual processing."

3.3.1 *Posner's orienting of attention: brain regions involved in shifting*

Some pathologies have provided insightful infor-mation through the study of attentional impairments associated with specific brain region lesions. Also, attentional processing has been studied in vivo on other mammalians as the macaque for instance Goodale and Milner

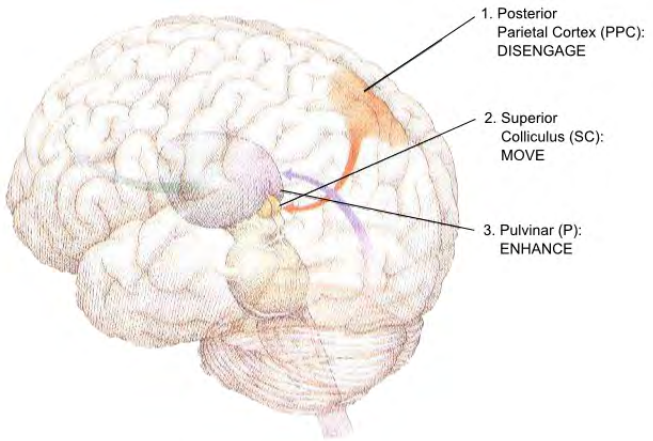


Figure 8: Posner triangular system of the orienting of attention. From (Posner, 1992).

(1992). Using a neuropsychological approach, Posner and Dehaene (1994) postulate that selective attentional processes are carried out by three distinct attentional networks: alerting, executive control and orienting. The "alerting'"network relates to sustained attention in response to a stimulus, the "executive control" network is engaged in activities that involve planning and decision making (Posner and Fan, 2008, Packwood et al., 2011), and the orienting network has been associated with disengaging, shifting, and re-engaging attention. According to Posner and Petersen (1989), this latter network is the result of an interaction of three *visual* areas: the Post Parietal Cortex (PPC), the Superior Colliculus (SC), and the Pulvinar (P) (see Figure 8). The PPC would be responsible of *disengaging* the focus of attention (FOA) in concert with the FEF that allow eye saccades to be triggered. Then the SC would reorient the FOA to the next target, using or processed visual infor-mation from the PPC or unprocessed directly from the retina. Finally the P would be responsible for enhancing the information at the FOA. The P is part of the thalamus which has been shown to enhance visual responses at the FOA via positive modulation of M- and P-layer responses in the LGN. This advocates for an early spotlight enhancement of the visual stimuli in the SC (Posner and Petersen, 1989). This models explains the neural circuitry of the orienting of attention. The next model presents the criteria that are used to determine which are the most important stimuli to be attended.

3.3.2 *Mental maps and feature integration: where to shift*

The most recent models of attention indeed include the notion of mental maps: a spatial map of the environment that would be used for storing the appropriate coordinates of a specific region of interest.

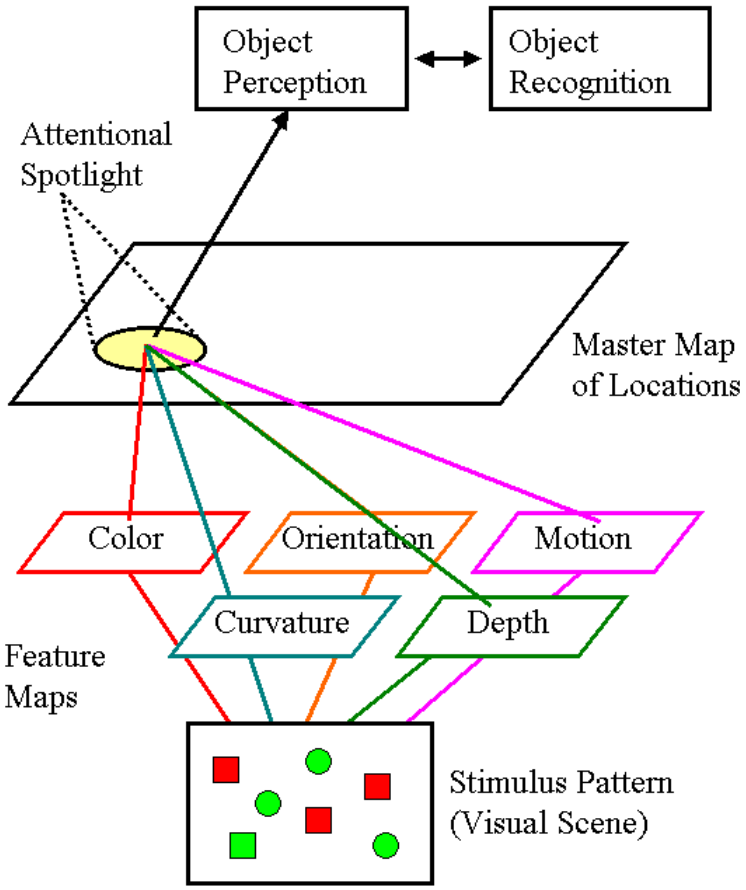


Figure 9: Feature Integration Theory model representation. From (Sutter, 2003)

But the concept of mental maps concern more than spatial features. They could also store independently other relevant features of the environment, like size, contrast, color, shape, etc.

Treisman and Gelade (1980) presented the first model using mental maps, introducing the master map of Locations storing"where" the object are in space. This model is called the "Feature Integration Theory" (FIT). It is a bottom-up model of parallel feature computation (as location, distance, color, orientation, size) that are stored in independent mental maps. These feature maps are then integrated in the master maps of Locations (see Figure 9). .

Michael et al. (2006) Michael and colleagues proposed a model-based on the Master Activation Map (MAM). The bottom-up feature integration aspect of the model is close to the Feature Integration Theory model (Treisman and Gelade, 1980). The top-down control is also taken into account and includes Kahneman (1973)'s limited resources

theory and an attention orienting system close to Posner and Petersen (1989)'s. The Master Activation Map would contain the location of the current most relevant stimulus according to salience, location and relevance to the goals (see Figure 10).

The interest of mental maps is to provide a tool that would store "where" to look next. Depending on the model, a competition between bottom-up saliency and top-down control (feature inhibition or amplification) is at stake. Wicken's SEEV give interesting insights on the bottom-up and top-down criteria competing for the attentional selection.

3.3.3 *Wickens' SEEV: to what target*

In a complex visual scene, stimuli compete for subsequent processing in higher visual centers (Lai et al., 2011). Wickens and McCarley (2007) introduced a model of selective attention that uses four criteria for electing the competition winner: Saliency Effort Expectancy Value (SEEV) :

- The *saliency* (S) of an element denotes its ability to capture the focus of attention (FOA). Extensive research has been carried out to propose algorithms for computing saliency maps in bottom-up approaches from visual features in images or videos as this explains to some extent the scanpaths in free viewing conditions (free viewing emphasizes bottom-up stimulation) (Filipe and Alexandre, 2013).

- According to Kahneman (1973), some stimuli require more *efforts* (Ef) than other to be processed. One of Kahneman's hypotheses is the fact that the maximum processing power is adaptable. Arousal for instance can temporarily extend this maximum, but also training, which automates some mental processes, or possibly the fact that multitasking requires different modalities. Wickens adapted this model and proposed the 4-D multiple resource model, which is also included in the SEEV (Wickens, 2008). For instance, eyes generally move alone for targets that present less than 20° of eccentricity, head is mobilized when the eccentricity exceeds 20°, and above 90°, the whole body rotates: the further away, the costlier it is to attend the target (Wickens et al., 2012).

- *expectation* (Ex) integrates the concepts of "How of attention" Gibson (1941) and "attentional set" Theeuwes (1994). It is the anticipation of a future stimuli resulting in the pre-conditioning of the brain through feedback connections. This allows for object search optimization when looking for a red object in an untidy room: the visual system will enhanced objects with red colour features.

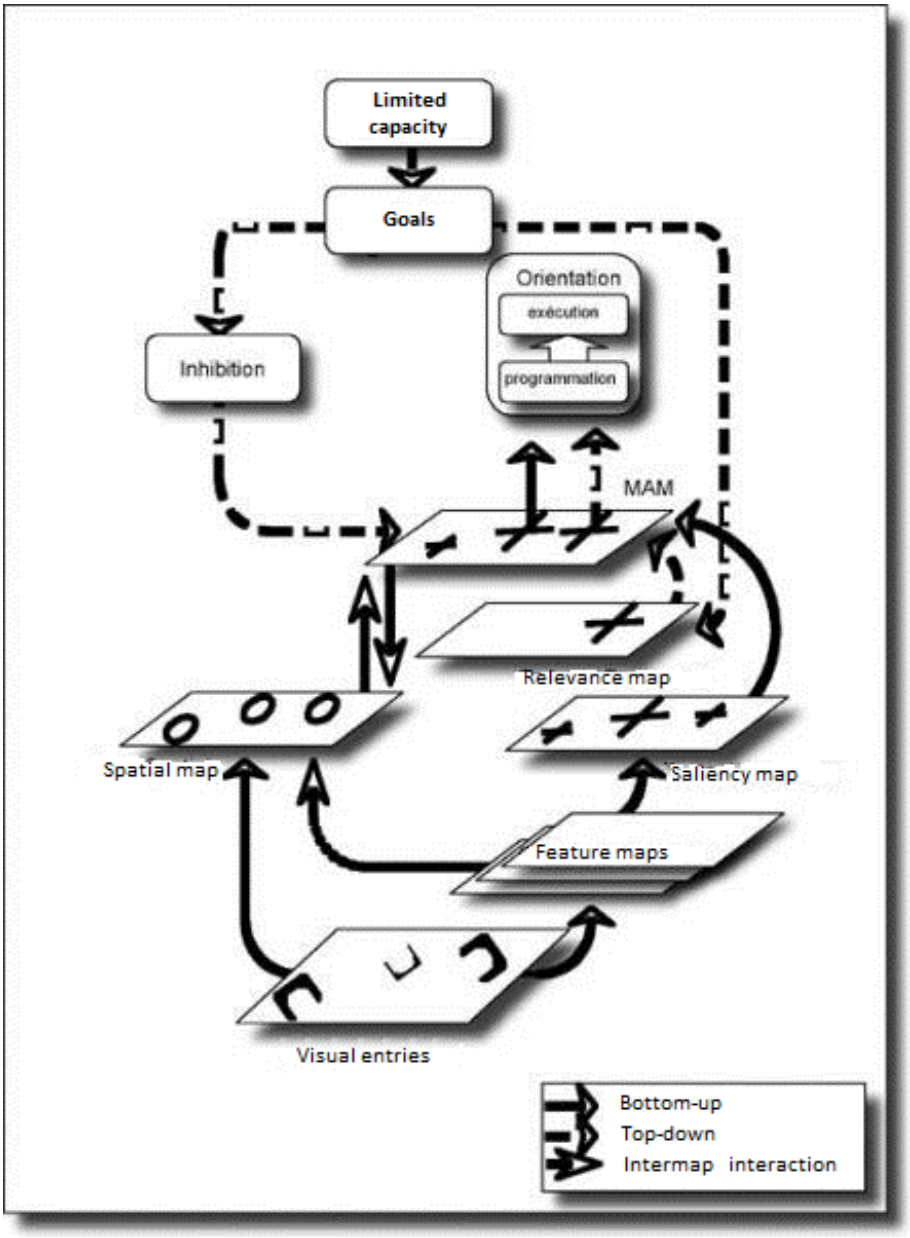


Figure 10: Master Activation Map model representation. From (Michael et al., 2006).

- Finally, *value* (V) is the estimated gain/outcome provided by using a stimulus rather than another one according to the task.

These components are then used for computing the probability of attending as the sum:

$$P(A) = S - Ef + Ex + V \tag{1}$$

The SEEV is a filtering model, therefore compatible with Posner's hypothesis of an early attentional selection. It associates Saliency and Effort as bottom-up features, as Expectations and Value are top-down computed features, but also includes a limited capacity modelling of attentional resources. Although this model explains "what" should be the next attended stimulus with a probabilistic approach, "when" to shift remains an open question. Findlay's modelling of attentional shifts is a proposition to deal with this issue.

3.3.4 *Findlay's competitive inhibition: when to shift*

Findlay and Walker (1999) proposed the "competitive inhibition" model that considers the *temporal* aspect of attentional shift (the "when"). In their model, attention is considered as being mainly an overt process, with very little space for covert shift of attention. It is called "active vision" because the brain starts to process peripheral information on the next target while preparing the eye orientation shift (Findlay and Gilchrist, 2001). This would allow for faster processing when the eye finally lands on the target. The shift between attending the same target for a longer time or move to the next one would be the result of a competitive inhibition continuously running between two structures :"where" to look at using features maps, and "when" to shift (see Figure 11).

The core of the model lies in the "fixate" versus "move" at Level 2. The fixate block is a gate (when the gate is open the eye is allowed to move at Level 1). The gate states depends on its inputs :

LEVEL 5: Voluntary control closing or releasing the gate.

LEVEL 4: Cognitive processing and temporal preparation impose constraints on the release of the fixation: either the cognitive processing, or the temporal preparation for the next saccade is not finished.

LEVEL 3: New visual fixation events can also advocate for more processing and block the gate to maintain the fixation (like a TV would do). But also peripheral visual events can require the eye to move immediately (like a tiger jumping at you).

The "where" side of the model is continuously updated so that the final "Level 2 move (salience map)" is able to compute Level 1 motor

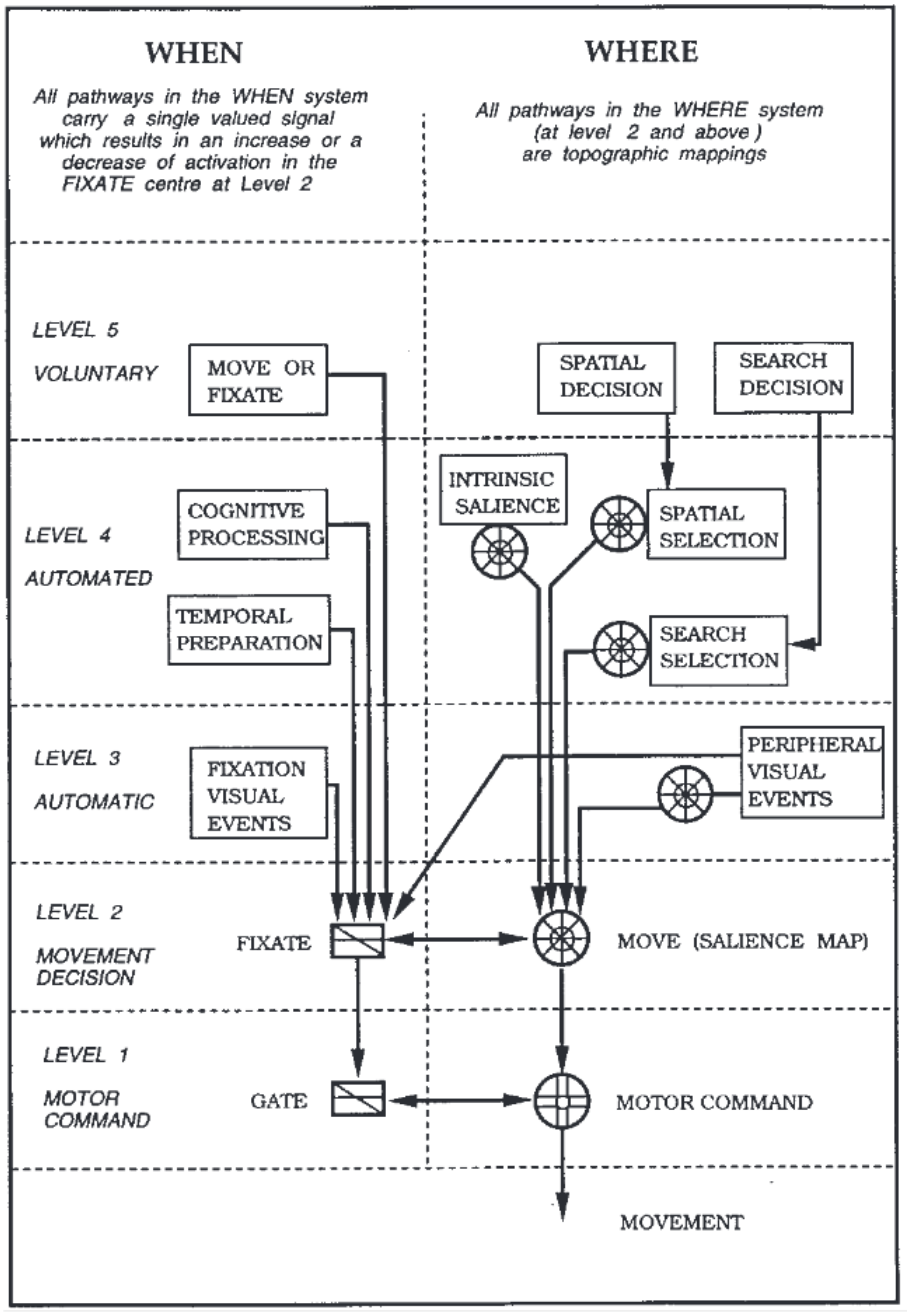


Figure 11: "When" / "Where" Competitive inhibition model representation. From (Findlay and Walker, 1999)

commands as soon as the "when" gate allows the eye to move. This results in an eye movement. The upper part of the "where" model includes intermediate maps whose values are due to bottom-up stimulation possibly moderated by level 5 top-down control. This model is compatible with Posner's model as it supports a sequential "what" analysis with a parallel "where" process. Also the 3 phases of the orienting of attention in Posner's model can be found in Findlay's model:

1. Attention would be *disengaged* according to the "when".
2. The "where" identifies the most relevant object for *orienting* the attention.
3. The "when" then controls *how long* the *engage* phase lasts.

3.3.5 *Conclusion*

Attention models are still very debated but a consensus about the type of features and how they are processed has emerged. Especially, the "where" and the "what" are processed independently, with the "where" processed in parallel, and the "what" sequentially, due to the limited capacity of the visual system. This is how the scene integration is possible. There are two forces that compete for the control of the eye and attention: bottom-up stimuli that attract the eye depending on their characteristics (for instance, a moving tiger in the jungle) and top-down control that can mitigate this attraction in order to favour other targets of interest (for instance, the rifle at your feet !). The relevant stimuli selection is based on features analysis. Features of the environment are computed and stored in coexistent mental maps regarding not only location but also colour, shape, size... Attention acts through feedback connections that moderates even early activations in the LGN therefore supporting the early filtering model of attention, though late selection is also possible. Two different streams of control are associated, the ventral and dorsal streams in the brain, that result in focal and ambient fixations. Focal fixations correspond to in-depth infor-mation encoding, whereas ambient fixations allow for a fast spread out scene integration.

Attention was first considered as an independent part of ourself, somehow immaterial, that could wander freely in the brain, would choose to focus at will, and would always precede vision. In recent models of attention, attention and vision intimately use the same structures. Attention shifts would be the triggered depending on levels of activation in the brain that are continuously updated by the vision outputs. Although some authors have highlighted the covert nature of attention, it is now accepted that attention is not only but almost always directed to where we are looking. This very strong

hypothesis is the reason for the study of eye orientation as a window into our mind, and justifies that we will focus on means of measuring ocular activity in the next chapter.

The presented models account for attentional processing in nominal situations, explaining what are the brain regions involved, which features are used for selection and what the attentional course in time. Nevertheless, attentional mechanisms are affected by the operational conditions that modify the operators' behaviours and decision making. These inter and intra-individual modifications are not explained by the models. Threfore the next sections introduce attentional impairments that have been reported as threatening the operators' mission success and dynamic attentional models that are useful for explaining the observed operators' behaviours, and metrics for measuring such impairments.

3.4 ATTENTIONAL IMPAIRMENTS

The attentional models presented in the previous sections are valuable for understanding the human infor-mation integration process. Our problem is to understand the underlying mechanisms of attentional impairment that lead human operators to neglect relevant information. We first propose to review the different concepts that have been put forward such as visual tunnelling, change blindness, inattentional blindness, attentional tunnelling. Possible factors and causes are then discussed to explain the occurrence of attentional impairments.

3.4.1 *Reduction of the useful field of view*

In automotive for instance, Crundall et al. (1999) reported that the omission of peripheral visual cues strongly contributes to inexperienced drivers' liability in car crashes. The phenomenon was also reported in aviation Williams (1995a), Thomas and Wickens (2004), and drone supervision (Waraich et al., 2013). This inability to detect peripheral hazard related cues was defined as a reduction of the useful field of view (UFOV), also known as "tunnel vision" (Williams, 1985). This reduction was explained through paradigms involving spatial presentation of peripheral stimuli. Indeed, it was shown that depending on the task, the detection was more adversely affected when the targets were placed at 2,2°, 7°or 10°than when placed at 0°(Plainis et al., 2001). A second explanation is related to task complexity. Indeed, tasks involving high perceptual load could leave little or no resources remaining for processing any additional infor-mation Lavie (1995). Williams (1985, 1995a), Macdonald and Lavie (2008), Plainis et al. (2001) have shown that increased foveal load and memory de-

mand shrink UFOV. Baddeley (1972) showed that high arousal and stressors (fatigue, operational pressure, stress) elicited a narrowing of the UFOV. Easterbrook (1959) made similar observations linking the occurrence of stressors and the reduction of the number of cues that could be attended.

3.4.2 *Inattentional and change blindness*

As previously stated, stress, perceptive and task load provide an interesting avenue to explain failure to notice critical cues in the environment. However recent experimental research proposed another approach. Mack and Rock (1998a), Newby and Rock (1998) revealed that non relevant and unexpected stimuli can remain unnoticed if presented in the foveal field and despite their saliency. They later defined this phenomenon as inattentional blindness. In their proof-of-principle experiment, Mack and Rock (1998a) asked participants to estimate which of the two bars of the cross presented in middle of a computer screen was longer. While the participants were processing the cross, an unexpected stimulus was presented *on top* of the cross for a duration as long as 200 ms *but remained unnoticed*. The phenomenon raised even more interest when Simons and Chabris (1999) demonstrated inattentional blindness in the context of *sustained* attention. Participants were asked to watch a video in which 2 teams (black and white) of 3 players are playing a ball game. Both teams are passing the ball to the other team members while running around in a small area. The participants had to count the passes performed by the white team. While participants are focused on this primary task, an unexpected actor dressed up as a gorilla enters the scene (see Figure 12), beats his chest, and leaves 6 seconds later. More than half of the 192 observers tested were *blind* to the gorilla due to the attentional focus on the white players, even-though the gorilla appeared on the screen for several seconds. According to Velichkovsky et al. (2002): "more than 50% collisions in road traffic arise from a missing or delayed hazard perception", in which drivers looked-but-failed-to-see (Mack and Rock, 1998a).

An adjacent concept to inattentional blindness is known as "change blindness" (for a complete review, see Simons and Rensink (2005)). For instance, Rensink et al. (1997) showed that participants failed to notice even large changes in successively presented pictures when a blank screen is inserted in-between two presentations (flicker), or when a physical object occludes the change (Simons and Levin, 1998). But the phenomenon is not only induced by those artificial interruptions but also by the ocular activity itself. Blinks and saccades indeed interrupt the visual contact with the environment. If the change is triggered during the interruption, it is likely to remain unnoticed (Kevin O'Regan et al., 2000). The phenomenon was reported in re-

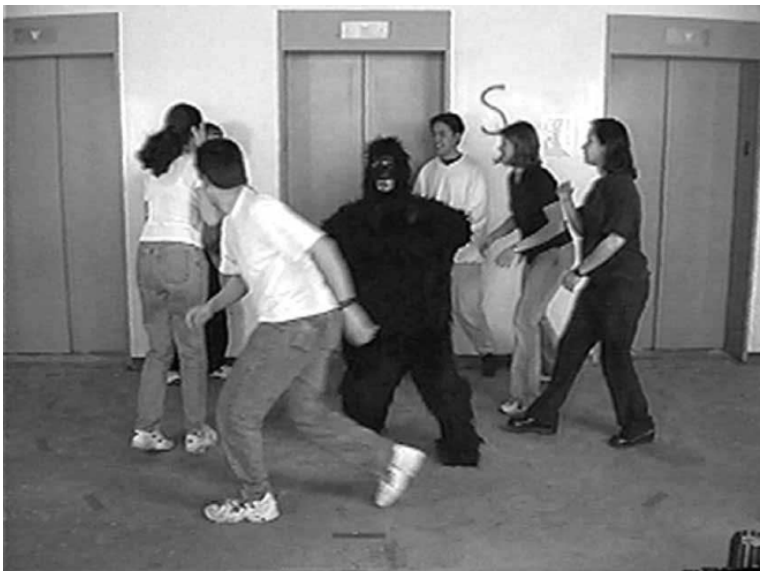


Figure 12: The Gorilla in our midst. Participants often fail to notice the unexpected gorilla, even when it stays in the midst of the screen and dances for a few seconds (Simons and Chabris, 1999).

alistic situations in automotive (Martens, 2011), drone supervision (Waraich et al., 2013) and aeronautics (Wickens, 2009) and pointed out as responsible for deadly accidents.

3.4.3 *Attentional tunnelling*

Inattentional or change blindness concepts may find some limitations as far as operational situations such as flying or driving are concerned. These two concepts refer to the inability to notice *unexpected/irrelevant* stimuli or changes in the visual scene. But it was reported that critical infor-mation is also missed in situations in which the operators *are aware and somehow expect* more important infor-mation on the other channels (Yeh et al., 1999). The explanation of such phenomenon requires the operational context to be taken into account, especially the fact that operators' decisions and actions are time-constrained. (Wickens, 2005) indeed formalized such impairment under the name of *attentional tunnelling*, which definition includes the time dimension:

> "the allocation of attention to a particular channel of information, diagnostic hypothesis or task goal, for a duration that is longer than optimal, given the expected cost of neglecting events on other channels, failing to consider other hypotheses, or failing to perform other tasks"

Studies carried out in flight simulators have shown that attentional tunnelling leads pilots to neglect critical pieces of infor-mation such

as imminent hazard cues (Wickens and Alexander, 2009), visual alarms (Dehais et al., 2003, Dehais, Causse, Régis, Menant, Labedan, Vachon and Tremblay, 2012) and auditory alarms (Dehais et al., 2014). As such, attentional tunnelling has been directly presented as responsible for nearly all the controlled flight into terrain (CFIT) accidents in a comprehensive study of 16,500 general aviation accidents between 1990 and 1998 (Shappell and Wiegman, 2003). Several explanations have to be put forward to explain this phenomenon.

First, it is worth noting that attentional tunneling occurs when human operators perform goals with high expectancy values such as reaching final destination (Dehais et al., 2014, Wickens and Alexander, 2009) or target identification (Dehais, Causse, Régis, Menant, Labedan, Vachon and Tremblay, 2012). In the SEEV model, high Value and high Expectancy visual stimuli (e.g. targets, landing ground) may promote excessive focus to the detriment of other elements of the environment. Goals with high expectancy are generally more demanding in terms of attentional resources (Dehais, Causse, Régis, Menant, Labedan, Vachon and Tremblay, 2012).

Such important and high level goals may affect the "when"/"where" competitive inhibition model in which the "fixate" gate would remain blocked (at Level 2) because of voluntary inhibition (at Level 5), high cognitive processing (at Level 4) or continuous new events at the locus of fixation (at Level 3) therefore overriding the calls from peripheral events (at Level 3) or directly from the "move" block (at Level 2). In other words, the operator would not be able to *move* his attention from the currently *fixated* target.

Eventually goals with high expectancy elicit high arousal (Dehais, Sisbot, Alami and Causse, 2011). Considering Posner's model, attentional tunneling could rely on a *temporary impairment* of brain regions such as the PPC due to *stress and high arousal* (Tracy et al., 2000). Indeed, there is evidence that impairment of the 'orienting' network (see Posner's model - previous section), induced by stressors (Pêcher et al., 2011), may cause attentional tunneling. Experiments conducted in a flight simulator have shown that the absence of response to either auditory or visual alarms may be explained by an inability to disengage attention: the warning systems are based on providing the operator with additional infor-mation, but this is of little use if the warning system is not also efficient at disengaging attention from the current task/target Dehais et al. (2003).

3.5 CONCLUSION

This chapter has presented what attention is, its main characteristics, and three comprehensive theoretical models of selective attention. They bring useful insights on the parameters influencing the course of attentional selection that is at stake in any of our actions,

but also in critical situations such as hazard detection, or the recovery of accident-prone situations.

The last section has added a new dimension to the attentional modelling issue. The intra-individual changes of the attentional selection and processing depend on endogenous and exogenous context. There is indeed a competition for attentional resources. But the maximum available resources might diminish because of cognitive load, perceptual load, stress, fatigue and conflict. In critical situations, the demand is higher than the maximum available resources. Operators might not be able to perform as usual, which result in critical stimuli to be omitted thus compromising the safety of the operations. As detailed, the consequences of the differences between what is expected from operators and what they can actually achieve is responsible for the deadly accidents in aviation.

It is therefore necessary to take these intra-individual changes into account while in operation. The goal is twofold. First, modelling them properly would result in better system designs that would be robust to operators' performance variability. Second, measuring attentional impairments while in operation would allow for piloting real-time interventions on the system.

There are many ways to measure objectively operators' attentional impairments. Indeed, Psychophysiology has proposed a wide range of physiological measurements that are related to psychological states. For instance, heart-rate has been elicited as a good stress indicator. Among all proposed metrics, eye movements appear to provide the most valuable metrics for their very tight link with attention. Eye-tracking is the tool for measuring the ocular activity, and was successful in reflecting attention since the early works of Yarbus in 1967. The next chapter presents the data flow collected by eye-trackers and the metrics that can be computed for detecting attentional tunnelling in real time.

4

EYE-TRACKING

RÉSUMÉ EN FRANÇAIS

Ce chapitre présente les différentes techniques d'oculomètrie, les données qu'elles génèrent, ainsi que les métriques oculaires qui en sont dérivées. Ces métriques couvrent un grand nombre de domaines d'applications, allant de la chirurgie, au neuro-marketing, en passant par l'étude des mécanismes attentionnels, notre sujet d'intérêt.

Parmi les différentes techniques d'oculomètrie présentées en introduction de ce chapitre, la vidéo-oculgraphie (VOG) est celle qui est le plus représentée de nos jours. Elle présente l'avantage de ne pas être intrusive, tout en proposant une très bonne résolution spatiale et temporelle. Elle consiste à analyser les images successives de l'oeil sous un éclairage infrarouge. Celui-ci apporte un bénéfice double: les réflexions des projecteurs infra-rouges sur la cornée sont des repères d'intérêt pour l'analyse de la position de l'oeil, et la pupille est magnifiée ce qui facilite son identification. A partir de ces deux repères, il est possible de calculer la position de l'oeil par le biais d'algorithmes qui associent une position de l'oeil à chaque image enregistrée.

Dans tous les cas, les oculomètres fournissent la position du regard, soit par rapport à l'environnement (VOG fixes), soit par rapport au visage du participant (VOG mobiles). Le traitement de la séquence de position de l'oeil est le même quel que soit le référentiel retenu: il s'agit d'identifier les événements oculaires (fixations, saccades et poursuite lisses). La méthode experte consiste à coder "à la main" les événements d'intérêt en inspectant visuellement les enregistrements. Cette méthode bénéficie des capacités humaines de reconnaissance de forme pour filtrer les artefacts de mesure, mais il est inconcevable de la pratiquer sur des enregistrements de longue durée. Pour cette raison et pour des raisons d'objectivité de l'analyse, la détection d'évènements oculaires est automatisée grâce des algorithmes spécifiques. Les 3 classes d' algorithmes les plus représentées s'appuient sur l'analyse de la dispersion des positions oculaires pour l'identification des fixations (I-DT), de la vélocité oculaire pour la détection des saccades (I-VT), ou sur la position relative du regard par rapport aux zones d'intérêt de l'environnement (I-AOI). Les limites de ces algorithmes sont pourtant bien connues: manque de robustesse face aux effets de seuil, grande diversité des implémentassions rencontrées dans la littérature limitant les possibilités de comparaison, manque de transparence de la part des constructeurs d'oculomètres qui de dévoilent pas leurs algorithmes d'identification et limitent les

degrés de liberté qui permettraient un contrôle fin de leur fonctionnement.

A partir de la séquence d'événements oculaires il est possible d'établir une grande variété de métriques oculaires, des plus basiques aux plus évoluées. Parmi les métriques basiques, on retrouve par exemple la durée, nombre et fréquence des fixations en fonction du temps ou des régions d'intérêt. Idem pour les saccades, les clignements. Les métriques liées à l'environnement (AOI-based) sont particulièrement pertinentes dans le cadre des facteurs humains: elles sont notamment corrélées avec l'intérêt relatif porté à chaque instrument du cockpit par exemple, ou avec la difficulté à en exploiter les infor-mations. Des liens direct entre ce métriques et la tunnélisation attentionnelle ont par ailleurs été montrés (Wickens, 2005). Elles nécessitent néanmoins de connaître la position de l'oeil dans l'espace, et la connaissance experte de l'espace environnant. A l'opposé les métriques directes ne nécessitent aucune connaissance experte de l'environnement mais ne permettent pas l'interprétation sémantique des positions du regard. On note par exemple le ratio fixation/saccade proposé par (Goldberg and Kotval, 1999), qui représente la balance entre la recherche d'information (exploration) et l'encodage (exploitation).

Plusieurs notions méritent alors d'être mises en perspective. Le chemin ventral permettant l'identification sémantique des objets (le "quoi"), semble correspondre au mode focal de la vision (fixations longues suivies de saccades de faible amplitude). Il permet donc l'encodage d'infor-mation ce qui se traduit par un ratio exploitation/-exploration élevé. Le chemin dorsal permettant la localisation des objets ainsi que l'étude de leurs mouvements, semble correspondre au mode ambiant de la vision (fixations courtes suivies de saccades de grande amplitude). Il permet donc l'exploration de l'environnement, ce qui se traduit par un ratio exploitation/exploration faible.

Fort de ces observations, Il est légitime de s'appuyer sur les métriques oculaires pour l'étude des mécanismes attentionnels en temps réel. Nous retiendrons deux approches distinctes qui font l'objet des deux contributions qui suivent ce chapitre: l'utilisation des métriques liées à l'environnement (AOI-based) qui permet une analyse spécifique en fonction du contexte de mesure et est particulièrement pertinent pour les facteurs humains, et l'utilisation de métriques directes ne requiert aucune connaissance de l'environnement et est plus générique.

La première contribution consiste en l'automatisation de l'analyse des métriques AOI-based pour la détection de la tunnélisation attentionnelle dans un contexte robotique. La deuxième contribution est donc une proposition de métriques directes, non dépendantes de l'interface, et calculables en temps réel. Pour leur calcul, un nouvel algorithme d'identification des états de l'oeil est proposé, l'Algorithme d'Identification des États de l'œil (ESIA dans le texte).

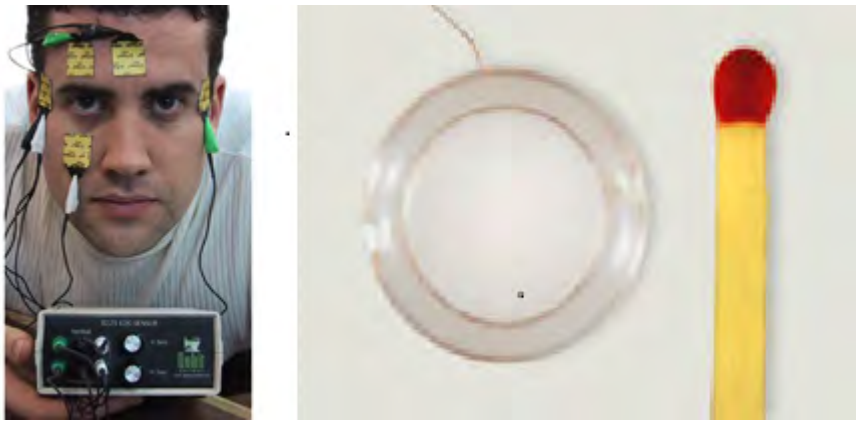


Figure 13: Eye-tracking systems. Left, EOG system from Qubit systems *http://www.qubitbiology.com* (2010). Right, Magnetic scleral search coil to be placed on the eye.

ENGLISH VERSION

Eye-tracking is now a mastered technology that has become affordable enough to have a large community of users. It is also more and more portable and embedded so that it is now possible to measure ocular activity in realistic conditions, possibly directly in operation. For an excellent review of eye-tracking systems, application, and methodology please refer to (Holmqvist et al., 2011, Duchowski, 2007).

4.1 EYE-TRACKING TECHNIQUES

There are four categories of eye-tracking techniques, which are presented in the following section.

4.1.1 *Electro-OculoGaphy (EOG)*

EOG is the measurement of the skin related potential around the eye. It is an indirect measurement for the oculomotor plant activity, which in turn is responsible for the eye position. This technique is still used for measuring the eye position, but also for filtering the EOG out of Electro-EncephaloGraphy (EEG) records. It requires electrodes to be placed around the eye, measuring only the relative eye-to-head position. An example of such a system is represented on Figure 13

4.1.2 *Scleral contact lens or search coils*

Another technique consists in placing an optical reference directly (a coil) on the eye which is then identified and used for computing the relative eye position, as for EOG. The coil can be replaced by a spe-

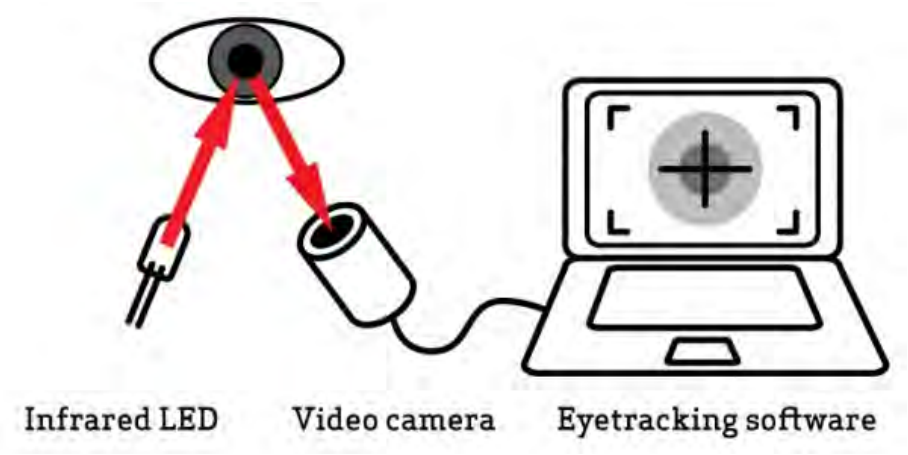


Figure 14: Simplified principle of VOG. From (Regis, 2011)

cific lens whose position is measured. Wired lenses are for instance detected via modification of the magnetic field surrounding the eye. This technique offers the best temporal and spatial resolution but are very intrusive as a foreign object has to placed on the eye.

4.1.3 *Video-OculoGraphy (VOG) and corneal reflection*

A video system records the image of the eye. Each image is analysed so as to compute the eye position from features detection like the pupil or the limbus (the boundary between the iris and the sclera). Used alone, VOG is limited to position the eye relatively to the head. A very interesting evolution of such a system adds a specific illumination of the eye (generally infra-red (IR) so that it is invisible), whose reflection on the cornea can be easily identified on video recordings (see Figure 14) . This technique has become the standard for recent eye-trackers. Corneal reflections form the four Purkinje reflections on the eye at the four dioptres between the exterior, the cornea, the anterior chamber, the lens, and the vitreous humour (see Figure 2 in Chapter 2).

Another very interesting characteristic of IR illumination is the *bright/dark pupil effect*. Because light entering the eye is absorbed in almost every direction, the pupil appears naturally black under IR illumination . If the IR light source is close enough from the camera central axis, pupil reflects IR light in such way that it appears bright on the video recordings facilitating the pupil identification. See Figure 15 for both bright and dark illumination configuration, and Figure 16 for the results on the video recordings.

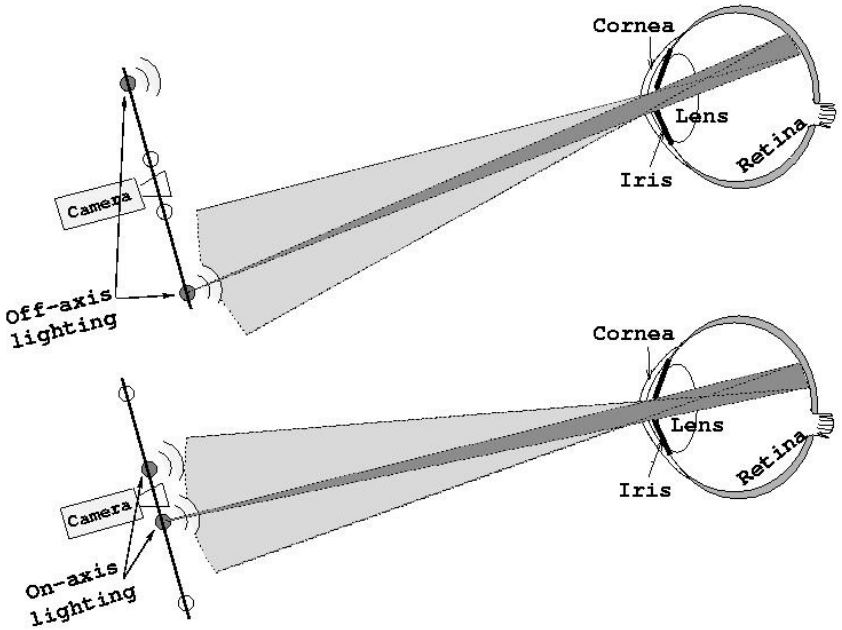


Figure 15: The two configurations for bright versus dark pupil effect on camera. (Meunier, 2009)

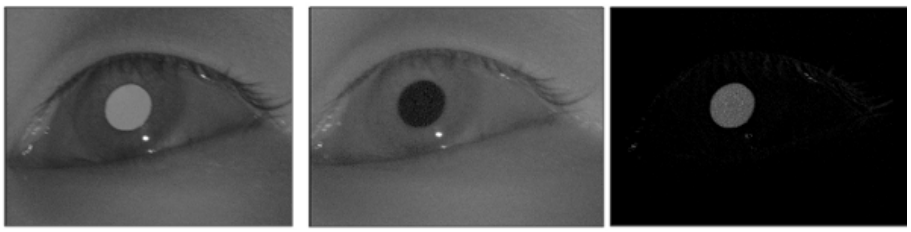


Figure 16: Bright versus dark pupil effect as seen on camera. (Meunier, 2009)

Figure 17: Eye-tracking systems. 1) Fixed and constrained, Eye-link 1000 Plus. The IR projector is the black grid on the right of the system 2) Fixed and remote on a desktop, SMI Red 500, both cameras and IR projectors are embedded under the screen 3) Fixed in a cockpit, Smart Eye Pro, IR flashes are the yellow dots close to the camera 4) Head-mounted, EyeTechSensor, IR flashes are on both sides of the eye camera

4.2 EYE-TRACKING SYSTEMS

Techniques have evolved since first eye-tracking systems from very intrusive to plug-and-play systems. This section only presents the main eye-tracking systems that are used today. They are all VOG. The three types detailed in this section are presented on Figure 17

4.2.1 *Fixed and constrained*

The most powerful eye-tracking systems are fixed and constrained. Fixed means that they cannot move in use, and constrained means that the participants that are required to have their head immobilized (usually a bite bar, or a chin rest). These systems are used for gaze monitoring in front of a perpendicular plan. The advantages of such a set-up is the high performance on both temporal and spatial resolutions. It is therefore suitable for the study of reading for instance, or gaze contingent experiments where stimulus on the screen change depending on the current eye movement. The Eye Link 1000 Plus is one example of such a system, sampling at 1000Hz. The eye position

is computed relatively to the stimuli area (or a projection screen, or a computer screen), in degrees or in pixels. See 1) on Figure 17.

4.2.2 *Fixed unconstrained*

This system has become the reference in neuromarketing and web usability studies as the eye-tracker does not interfere with the user and allows for on screen localization of the gaze simultaneously with the recording of the screen contents for further analysis. The head-box (volume in which the head is allowed to move while measuring) is quite small. Sampling is relatively fast, up to 500Hz, with a good spatial resolution. It is generally limited to computer related contents with eye coordinates computed in pixels. See 2) on Figure 17.

4.2.3 *Remote in situ*

Other remote eye trackers can be set-up in specific environments of interest, as a cockpits for instance. They require the 3D virtual modelling of the environment, which is quite tedious. The head-box of such systems is larger and the combination of more cameras (up to 8 with Smart Eye Pro), allows for all-around measurement in wide cockpits. In planes for instance, buttons are situated not only on the front panels but also on the ceiling and on both sides of the pilots, requiring the head to rotate freely to be accessed. Depending on the system, the head tracking can be performed from natural face markers, or from an external stylus that is fixed on the participant's head. These systems sample up to 250Hz and have a moderate precision. This can threaten the context related analysis of the eye-tracking data as it might be hard to discriminate where the eye is exactly pointing between two instruments for instance. The eye position is generally computed in degrees relatively to a reference axis arbitrarily chosen in the virtual environment. See 3) on Figure 17.

4.2.4 *Head-mounted eye-trackers*

Head-mounted eye trackers are very relevant in free moving environment as a supermarkets, but also in ecological environments as cockpits. Their sampling is generally slower, around 100Hz. The tracking of such systems is performed in two independent steps. First the eye-in-head position in degrees is computed from the eye camera, which is placed under the eye. Second the head-in-environment position is computed from the field camera that is installed on the eye-tracker mount and pointing forward. This finally allows for the position of the eye in the environment to be computed from the composition of both movements. See 4) on Figure 17.

4.2.5 *Comparison*

The previous sections give an overview of different set-ups that can be used for tracking the eye. Each set-up comes with weaknesses and advantages that have to be evaluated in the light of the measurement context. The more constrained, the better performance, whereas the less intrusive ones allow for ecological measurement at the price of reduced performance.

Another critical aspect of eye tracking is sampling frequency. The Shannon theorem shows that sampling determines which phenomena that you can be captured. Eye tracking systems have sampling frequencies ranging from 10Hz to 2000Hz. Sampling must be compared to the type of eye movements that are to be measured.

Another main struggle with eye-tracking comes from gaze analysis. This is required when the analysis aims at comparing the use of the different instruments of an interface for instance. With fixed systems as the SMI Red, the geometrical configuration is fixed, which facilitates the operation. But in all other cases, a model of the world has to be provided. This is a very costly step in the process of eye-tracking, time-wise and precision-wise which adds interventions in the measurement process that are hard to control from a quality point of view. Using head-mounted systems mitigates this problem as the field camera provides the point of gaze directly after a simple calibration of the eye camera. If context related gaze infor-mation is to be analysed, identifying the objects in the field camera view implies putting markers in the environment. Then, a post processing is required for repositioning the world according to these markers, and finally identifying which elements in the world were gazed at.

4.3 EVENT DETECTION

Eye-trackers provide the eye positions. Nevertheless, analysing the sequence of eye positions alone does not reap up the full benefits of eye-tracking. Because of the nature of the eye movements (see Chapter 2), eye-tracking data can also be analysed as a sequence of fixations and saccades (Holmqvist et al., 2011). An example of an eye-tracking record of a "fixation-saccade-fixation" sequence is presented on Figure 18. This event detection is almost systematically performed when using eye-tracking, as it relates the functioning of the visual system, and therefore attention (see Chapter 3).

A critical parameter of eye-tracking event detection is the sampling frequency which impacts ocular parameters as eye velocity. Figure 18 is a good example of such a phenomenon: the 1000Hz sampling in blue relates the saccade speed correctly, whereas the 25Hz sampling fails to model the right slope because of subsampling (Shannon, 1949).

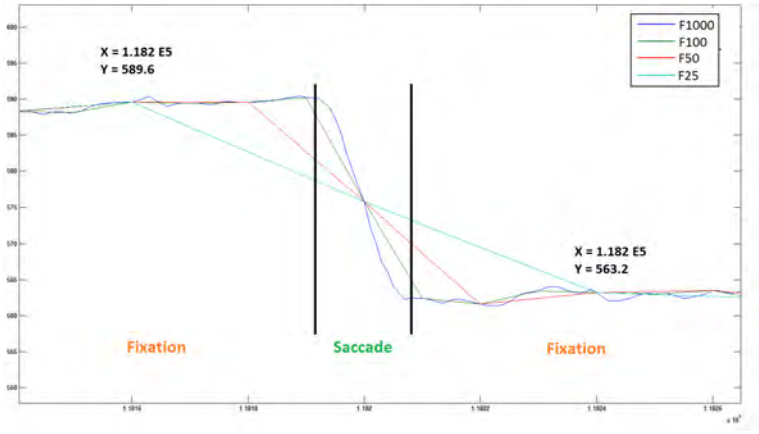


Figure 18: Record of the vertical position of the point-of-gaze. This record is equivalent to a fixation-saccade-fixation sequence. The different plots correspond to simultaneous sampling at different frequencies. The dispersion (vertical position) is not affected, but velocity (slope) is.

The different methodologies for extracting events from eye-tracking records are presented in this section.

4.3.1 *Manual coding: expert*

Manual coding is the identification of the ocular events performed by human experts on the basis of the eye positions series. This method is *very time-consuming*, and exposed to human subjectivity and inconsistency, but benefits from the excellent pattern matching ability of human experts (Holmqvist et al., 2011).

This technique is of particular interest and might be the only one reliable when identifying fixations in a gaze-overlaid video i. e. when the analysed record is the field camera video, with the gaze direction superimposed. This allows for content related analysis in highly dynamical environment (Tatler et al., 2005).

Indeed Munn et al. (2008) tested three coders and an automatic algorithm on the same event detection task and showed that the human coders agreed more often with the other coders analysis than with the algorithm. Try it yourself: any reader of this manuscript will easily identify the two fixations in Figure 18.

4.3.2 *Dispersion based (I-DT)*

Dispersion-based algorithms are meant for detecting fixations as being sequence of eye positions that are spatially grouped together in an predetermined area. It can be a circle or a square, whose size ranges from 0.5° to 2°, 1° being the most used value (Holmqvist et al., 2011),

which corresponds to an average of 30 pixels. The algorithm is very simple, the dispersion threshold allows for a straightforward control, and works in real time. It is also acknowledged as being the most reliable fixation detection algorithm according to Salvucci and Goldberg (2000). It is especially relevant with low frequency sampling as dispersion is hardly affected by sampling contrary to speed, as illustrated on Figure 18. At higher frequencies than 100Hz, saccade speed is correctly sampled which advocates for velocity based algorithms (Holmqvist et al., 2011).

4.3.3 *Velocity based I-VT*

Velocity based algorithms are meant for detecting saccades as periods of time during which the velocity is above a predetermined threshold. This threshold can be of different natures for the onset (start) and the offset (end) of the saccade. Thresholds in the literature typically range from $20° s^{-1}$to $130° s^{-1}$(Holmqvist et al., 2011) and depend on the application. Small extent saccades (few degrees) have a typical speed around $50° s^{-1}$whereas high extent saccades (several degrees) reach velocities above $600° s^{-1}$. For instance, Duchowski (2007) declared that a threshold of $130° s^{-1}$ "should effectively detect saccades of amplitude roughly larger than 3°". The weakness of velocity based algorithms is their sensitivity to noise and to the sampling frequency. Noise indeed modifies the position of consecutive samples, therefore impacting velocity, as it is computed from the distance between two consecutive samples divided by the sampling period. The noise amplitude being fixed, the higher the frequency, the higher the velocity noise.

When the saccade speed is well sampled (with a high frequency sampling when not noisy, or possibly filtered), the maximum saccade speed can be computed. This value is then used to reject non physical saccades that reach speeds typically above $750° s^{-1}$(Duchowski, 2007).

Derived from velocity based algorithms, acceleration based algorithms are also used, but the noise is amplified with a multiplication factor (due to the double derivation from eye position series), which makes them moderately reliable methods, which are used mainly for online segregation between smooth pursuit and small saccades (Holmqvist et al., 2011).

4.3.4 *Area-based algorithms I-AOI*

In most studies, the focus is on context-related analysis through the definition of Areas Of Interest (AOIs). AOI-based algorithms compare point-of-gaze positions in the environment relatively to the AOIs (see Figure 19). As long as the gaze "belongs" to an AOI then it is considered as a "hit". The period during which the eye remains in the AOI

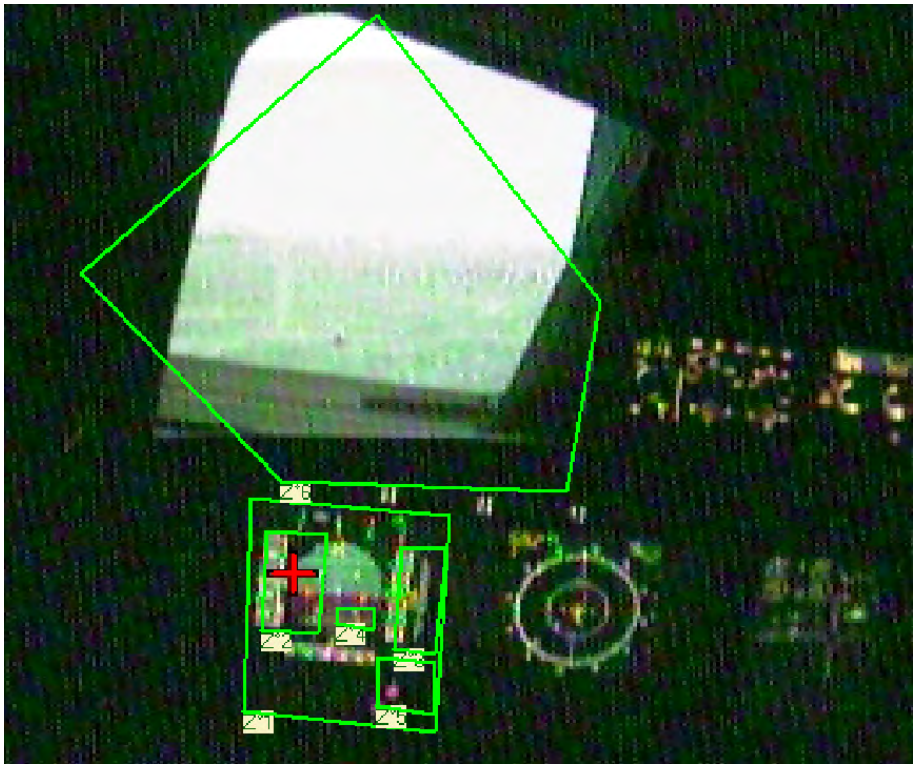


Figure 19: Areas of interest, in the context of a flight simulation. Each green polygon is an AOI. AOIs can overlap. EyeTechSensor

is called a the *dwell* time. This analysis is particularly relevant when a comparison between the use of different instruments has to be made (order of importance, frequency of consultation, AOI scan paths).

As explained in Table 7, AOIs require the environment to be modelled and the head motion to be compensated for point-of-gaze calculation. It is therefore a costly analysis. It should also be noticed that adding computation steps in the analysis weakens the signal/noise ratio because of the errors that are introduced at each calculation steps of the environment repositioning. Recordings having good eye-in-head data might finally result in non usable eye-in-environment data because of the added noise.

4.3.5 *Probability based I-HMM*

Statistical methods are also used for identifying fixations and saccades from the velocity distribution. Salvucci and Goldberg (2000) proposed a two-states Hidden Markov Model: the fixation state would emit speed distributed around slow velocities ($5° s^{-1}$) whereas saccades would emit speeds around higher velocities ($400° s^{-1}$). Though this method is elegant, it is harder to control than I-VT and I-DT,

which are the most implemented event detection algorithms in commercial eye-tracking software.

4.3.6 *Tree-based algorithms I-MST*

Another interesting method, which is not available in real time but post-hoc only, is based on Minimum Spanning Trees (MST) (Salvucci and Goldberg, 2000). This method is not discussed here as it is not used much in the literature. It indeed requires acute knowledge of MST to appropriately control the criterion separating fixations that are expressed in terms of edges length and depth.

4.3.7 *Blinks detection*

Blinks detection is rarely the base of a specific detection algorithm, though it should. Blinks are generally detected as the absence of data, due to pupil loss (the pupil is hidden during the blink) (Holmqvist et al., 2011). Nevertheless some signatures have been proposed, such as a pupil diameter reduction to zero followed by an increase. But this measurement is not reliable as the eye lashes cross the pupil detection zone during a blink and, because they are generally dark, might be detected as part of the pupil by the pupil detection algorithm. Also, pupil non-detection might come from other sources that are not linked with blinks, because of temporary bad illumination for instance.

4.3.8 *Conclusion*

Holmqvist et al. (2011) proposed a nice conclusion to this chapter, highlighting the need for better algorithms to be developed:

> None of the algorithms described in this chapter detect and measure all events reliably. They all have settings that are not easy to grasp for the beginning eye-tracking researcher, but which have profound effects on the results produced. Most worryingly, the algorithms treat the raw samples so differently that basic measures such as fixation duration and saccade amplitude will be difficult to compare between algorithms using the same data. The majority of eye-tracking researchers appear to have faith in the algorithms provided by manufacturers, and manufacturers mostly support this faith when they hide the algorithm properties in settings dialogue that are difficult to see through, and when not simultaneously showing raw data and detection results in the same scan path or velocity graph.

This important point played a major role in deciding to propose a new identificaiton algorithm in this Ph.D.

4.4 OCULAR METRICS

The event identification produces compressed ocular data under the shape of a sequence of events. Depending on the type of study that is carried out, different types of events are used. In our domain of application involving realistic conditions experiments, measurements are noisy because of the uncontrolled context, with slow sampling (50Hz) due to the use of head-mounted systems. This restricts the type of events that can be detected: fixation and saccades can still be identified by dispersion based and velocity based algorithms. One specificity of such conditions is the omnipresence of smooth pursuits that are due to the motion of targets relatively to the head, but also due to the head motion relatively to the environment (which is not the case in constrained set-ups).

The following section therefore focuses on metrics that are computed from saccade-fixation sequences only, discarding other small extent phenomena as tremors, micro-saccades, drifts, vergences, glissades. Though it is a limitation of the infor-mation that could be extracted from the ocular activity, the saccade-fixation sequences contain the most used infor-mation for ocular activity analysis. Indeed most of the eye-tracking software, even the ones sampling at high frequency with high signal on noise ratio limit their identification to fixation and saccade events (Holmqvist et al., 2011).

4.4.1 *Basic metrics*

Simple metrics are generally computed from the saccade-fixation sequences, as mean fixations duration, number of fixations, saccades extent, and frequency, and dwells hits (periods in a specific AOI, eventually several fixations and saccades in one dwell). This subsection presents the most used basic metrics and their interpretations.

FIXATION: Number of fixations per AOI (Poole et al., 2005), Mean fixation duration pr AOI (Svensson et al., 1997), Ratio of the total duration on an AOI on the total time of the experiment (Flemisch, 2000).

Interpretation: According to Goldberg and Kotval (1999) the more fixations on a given task, the less efficient the search. A large number of fixations in a AOI denotes the fact that this AOI is more important and salient than others (Poole et al., 2005). More frequent looks at an AOI compared to others would notify that this AOI is of particular interest regarding the task (Jacob and Karn, 2003). Longer fixations would also relate a higher diffi-

culty to extract infor-mation and a higher propensity to require attentional processing (Just and Carpenter, 1976, Goldberg and Kotval, 1999).

DWELL: Total number of dwells, Dwell number on each AOI, Mean dwell duration on each AOI (Albert and Liu, 2003). Dwell time on an AOI (Hauland, 2003), Long dwell frequency on each AOI (Svensson et al., 1997), Dwell distribution (Mello-Thoms et al., 2002), Number of AOI used (Thomas and Wickens, 2004).

Interpretation: Wickens (2005), Mello-Thoms et al. (2002) postulate that the dwell rate distribution over all AOIs relates the relative importance and the attention allocated per AOI. For Fitts et al. (2005), the frequency is the importance and interest indicator. Moreover the number of AOIs consulted diminishes with attentional tunnelling (Thomas and Wickens, 2004).

SACCADES: Number and frequency of saccades Goldberg and Kotval (1999), Peak velocity (Di Stasi, Marchitto, Antolí and Cañas, 2013) and main sequence (Bahill et al., 1975, Di Stasi et al., 2011, 2012).

Interpretation Saccades do not relate to infor-mation extraction therefore do not provide infor-mation on the saliency or the complexity of an AOI. But a high number of saccades relates an active search Goldberg and Kotval (1999). Attentional focus is also linked with a reduction of the saccadic activity Tsai et al. (2007), Cowen et al. (2002). The saccadic velocity was also used to estimate mental workload Di Stasi et al. (2011), drivers' fatigue Di Stasi et al. (2012), and arousal Di Stasi, Catena, Cañas, Macknik and Martinez-Conde (2013).

BLINKS: Frequency of blinks (Bruneau et al., 2002), Time between two blinks (Veltman and Gaillard, 1996).

Interpretation A lower blink frequency would denote a higher workload according to Wilson (2002) and an important visual load Veltman and Gaillard (1996) whereas an increase in frequency would be linked with fatigue according to Bruneau et al. (2002), Brookings et al. (1996). But blink frequency and duration also adapt to ambient light Goldberg and Wichansky (2003) which imposes the ambient light to be controlled (still this does not compensate for variations in the luminosity of the stimuli).

PUPIL Size and variation of the pupil (Brookings et al., 1996, Goldberg and Wichansky, 2003). Index of Cognitive Activity Marshall (2002).

Interpretation A bigger pupil size relates more cognitive efforts according to Brookings et al. (1996). But this parameter is also

very sensitive to variations in light level (Goldberg and Wichansky, 2003). Nevertheless, according to Marshall (2002), the analysis of high frequency variations of the pupil diameter would allow for cognitive activity analysis discarding the changes in light level.

FIXATION/SACCADE RATIO: Goldberg and Kotval (1999) defined the as the time spent in Fixation divided by the time spent in saccade during an experiment. It is therefore an agregation of two basic metrics. They used this ratio for comparing interfaces.

Interpretation High ratios were associated with interfaces that required more processing or less ocular search than interface with lower ratio. It is an interface-independent measure that is of special interest for our purpose.

4.4.2 *Heat maps*

Another metric that is extensively used in neuro-marketing is the heat map. Heat maps represent the accumulation of successive eye positions on the reference image so that the contents that attracted the attention of the user is straightforwardly represented.

There are different ways of calculating heat maps. Either they use raw data, i. e. the sequence of eye positions; or they use the fixations identified in the raw data. Each of the N source points (resp. each eye position, or each fixation) is then used to compute a 2D Gaussian distribution covering the span of the reference image with the following formula :

$$\forall i \in [\![1, N]\!], \quad G_i(x, y) = \exp\left(\frac{-(x - x_i)^2 - (y - y_i)^2}{2\sigma^2}\right) \tag{2}$$

where $G_i(x, y)$ is the value of the Gaussian distribution at (x, y) coordinates, (xi, yi) are the source points coordinates, and σ^2 is the *variance* of the distribution.

This calculation is performed for all source points. The accumulation map is then the sum of all Gaussian distributions defined in Equation 2, that is then normalized :

$$A(x, y) = \sum_{i=1}^{n} G_i(x, y) \tag{3}$$

$$A_{normalized} = \frac{A(x, y)}{\max_{x,y}(A(x, y))} \tag{4}$$

For each pixel in the reference image, a colour (or a transparency level) is associated with the value on the scale between 0 and 1. This coloured (transparency) map is then superimposed on the reference

Figure 20: Heat map calculated from eye-tracking data recorded during a period of manual flying in a flight simulator. It highlights the use of the Out Of the Window (OTW) and the speed and altitude indicators on the Primary Flight Display (PFD)

image, which produces the *heat map* or *attention map* (Holmqvist et al., 2011). Areas of the map with high accumulation are pictured in red, whereas unattended areas remain black (see Figure 20).

This representation has become popular because it is visually convenient and facilitates communication on eye-tracking data analysis. Indeed, heat maps of two groups of participants can easily be compared but require the reference image to be the same. Despite their practical use, heat maps are quite limited: they do not account for the transitions between the different elements of the environment and must be computed a posteriori for normalization purpose (Holmqvist et al., 2011).

4.4.3 *Scan paths*

In order to account for the saccadic activity, one can study the sequence of saccades and fixations in the environment. Scan paths, first introduced by Yarbus et al. (1967) (see Chapter 2), provide valuable infor-mation. Not only do they show the positions that were occupied by the eye, but also the transitions between the different elements of the environment. In the usability domain, scan paths are extensively used as they relate the "cognitive" path followed by the operator when performing specific operations. This allows for an in-depth usability evaluation (Poole et al., 2005, Jacob and Karn, 2003),

Figure 21: Basic-T scan path. From (*Basic-T scan path*, n.d.)

revealing shortcuts that operators take, which would be impossible to detect by asking them directly.

When combined with the AOIs, scan paths can be summarized as AOI sequences. These sequences are of great interest as they relate the repetitive use of a group of instruments for instance, with a specific pattern. In flight simulations, pilots scannig routines appear like the "basic-T" scan (see Figure 21). For methods for comparing scan paths, please refer to (Le Meur and Baccino, 2013). The next subsection details tools for analysing AOI related measurements.

4.4.4 *AOI Sequence, Transition Matrix and Markov Models*

Sequences of AOI hits can easily be represented as strings. For instance, a pattern involving AOI B, C, D, then A, will be represented as BCDA. In the course of an experiment, these strings might be very long. A simple way to analyse them is to restrict the length of the studied strings to 2 or 3 for instance. All strings of this length are then isolated: for a length of 2 isolated in the string taken as example, this would result in BC, CD and DA. Then statistics can be computed for finding the most frequent string of length 2 via counting the occurrence number of each 2-letter string.

Another common way of analysing AOI related transitions is *transition matrices* in which the number of switches from any AOI to another is represented. Holmqvist et al. (2011) report that most of the AOI sequences studied are of length 2, whereas some authors demonstrate that sequences longer than 2 are generally not repeated across participants, compared to direct transitions from an AOI to another (Duchowski, 2007).

Finally, Markov chains are also deployed for their probabilistic approach to AOI transitions. In a first-order Markov chain, the probability of transiting from an AOI to another is represented on the edge joining the two places labeled with the AOI names. Again, higher order Markov chains can model the probability of sequences longer than 2 but these are hardly used (Holmqvist et al., 2011). An example of first-order Markov chain was proposed for off nominal activity detection in supervision (Hayashi et al., 2005).

4.4.5 *Ambient versus focal*

An interesting approach for studying the ocular activity is combining each fixation with the following saccade Unema et al. (2005), Velichkovsky et al. (2000, 2002, 2005), Follet et al. (2011). Velichkovsky et al. (2005) indeed elicit a systematic combination, which can be classified according to the fixation duration. Duration belonging to the 90 ms - 260 ms segment are followed by saccades of amplitude greater than 5°, which corresponds to the periphery of the visual eye field with poor visual acuity. They call these fixations *ambient fixations*. Oppositely, fixations of duration longer than 260 ms-280 ms are associated with smaller extent saccades falling in the parafoveal region of the retina (less than 5°), they are called focal fixations. This model supports the two visual pathways scheme from Goodale and Milner (1992). Identifying focal versus ambient fixations in the ocular activity therefore relates the visual attention activity. Velichkovsky et al. (2002) propose an express diagnosis of hazard detection: according to their study, the focal versus ambient fixations could differentiate participants who detected a potentially hazardous targets and those who did not. This study supports the use of eye-tracking for improving safety in automotive as previous studies did (Chapman and Underwood, 1998). It especially advocates for *real-time and short period based analysis of eye tracking data when applied to domains involving dynamic scenes*.

4.4.6 *Eye-Fixation-Related Potentials*

Focal versus ambient fixation is an interesting fixation classification link with cognition. Classifying fixation depending on their cognitive role can also be done thanks to multi-modality. One very promising avenue has been proposed by Baccino and Manunta (2005). They coupled an electroencephalograph (EEG° and an eye-tracker in order to identify the neural activity associated with ocular fixations. This technique called Eye-Fixation-Related Potentials (EFRPs) was successful in the investigation of parafoveal processing in reading: EFRPs could differentiate pairs of words, whether they were associated words, nonassociated words or nonwords (Baccino and Manunta, 2005). A

more recent study demonstrated that this technique was a useful tool to study temporal dynamics of visual perception and processes underlying object identification (RäMä and Baccino, 2010).

4.5 CONCLUSION

Eye-tracking provides very rich data that result in various metrics. The different steps consist in finding the pupil position in the eye then computing its position in the environment, using this sequence of eye positions to identify ocular events as fixations and saccades, and finally computing ocular metrics.

Ocular metrics can be basic as the mean fixation time, or the frequency of saccades, but also more sophisticated, as heat maps or transition matrices. These metrics have been proven meaningful in various usability studies but also in realistic simulations. They relate the use of the interface and the importance given preferably to certain instruments against others. Most interesting for our purpose are the metrics that are directly linked with attentional phenomenon. Especially the transitions between AOIs correspond to the shifts between infor-mation sources. Oppositely, long fixations should be associated with a focus of the attention on a specific source of infor-mation. One natural hypothesis is therefore that such metrics would be sensitive to attentional impairments like attentional tunnelling.

In order to test such an hypothesis, a specific robotic experiment was designed for inducing attentional tunneling. In the next chapter, the metrics of attentional tunneling presented in this chapter were not only calculated and evaluated for their sentivity to attentional tunneling, but an automated attentional detector was also proposed.

5

FORMAL DETECTION OF ATTENTIONAL TUNNELING IN HUMAN OPERATOR-AUTOMATION INTERACTION

RÉSUMÉ EN FRANÇAIS

Le but de cette thèse est de contribuer à la détection en temps réel des états attentionnels d'un opérateur pour limiter les risques d'accidents en contexte d'interaction homme-machine adaptative et pour donner des outils aux experts en facteurs humains. Dans certaines conditions, la tunnélisation atttentionnelle survient et perturbe les capacités de prise de décision en provoquant une focalisation excessive de l'attention. Nous avons également montré que certaines métriques oculaires sont sensibles à ces modifications des mécanismes attentionnels. Deux approches ont été identifiées au chapitre précédent, l'une s'appuyant sur les métriques AOI-based, et l'autre sur les métriques directes. Ce chapitre présente la première contribution de cette thèse qui consiste en la vérification de la pertinence des métriques AOI-based pour la détection de la tunnélisation attentionnelle dans un contexte robotique, et en l'automatisation de leur analyse grâce à une méthode d'apprentissage automatique. Cette étude est organisée en plusieurs étapes qui sont détaillées dans le paragraphe suivant. Elle a fait l'objet d'une publication dans IEEE Transactions in Human-Machine Systems (Regis et al., 2014).

Dans un premier temps, l'expérience robotique est décrite ainsi que les principaux résultats comportementaux. Les 23 participants mobilisés pour cette étude ont eu pour mission de superviser un robot à distance depuis un ordinateur de bureau. Le robot est équipé de plusieurs capteurs (GPS, ultrason, vidéo panoramique) dont les informations sont conditionnées puis présentées sur l'interface de l'ordinateur. La vidéo panoramique est notamment affichée sous forme d'un bandeau vidéo dont l'interprétation est spécifiquement rendue difficile. Pendant la mission, une alarme est présentée aux participants alors qu'ils sont concentré sur le pilotage du robot. En fonction de leur réaction à l'alarme, les participants sont classés en trois groupes. Le groupe TUN=0 (n=10) est constitué des participants qui détectent l'alarme dès son apparition et suivent les consignes associées. Le groupe TUN=1 (n=8) est constitué des participants n'ayant pas pris en compte l'alarme pendant au moins 50 secondes. Le groupe TUN=0/1 (n=4) est constitué des personnes qui ne détectent pas l'alarme immédiatement, mais la prenne finalement en compte avant 50 secondes de présentation et finalement suivent les consignes asso-

ciées. Les participants sont équipés d'un oculomètre et d'un électrocardiogramme dont les données, permettent le calcul de deux métriques oculaires spécifiques de la tunnélisation attentionnelle: nombre d'AOIs consultées sur 20 secondes (NBAOI) et taux de transitions entre AOIs (SWR) et une métrique de stress (HR). Il est alors vérifié statistiquement que ces trois métriques sont bien des marqueurs de la tunnélisation attentionnelle par comparaison de leurs valeurs moyenne en fonction du groupe de participants. Fort de ce résultat, il est envisagé d'entraîner un classifieur pour automatiser l'identification des instants de tunnélisation attentionnelle pendant la mission. Pour des raisons de robustesse aux entrées continues et bruitées, le classifieur retenu est un Système d'Inférence Flou Neural Adaptatif (ANFIS). Ce système présente par ailleurs l'intérêt de pouvoir être traduit en langage naturel, ce qui facilite son analyse par les experts en facteurs humains. Les différentes étapes d'entraînement de l'ANFIS sont alors détaillées, de la construction de la base d'apprentissage, à l'évaluation de la performance de classification, en passant par la supervision de l'apprentissage par cross-validation. Les performances de l'ANFIS sont ensuite comparées avec un classifieur classique (une Machine à Vecteur de Support, ou SVM), où l'ANFIS se montre légèrement meilleur. Finalement, les lois d'apprentissage de l'ANFIS sont traduites en langage naturel et comparées aux lois utilisées par les experts pour la détection de la tunnélisation attentionnelle. Il apparaît que celles-ci sont remarquablement cohérentes. Pour finir, il est montré que le classifieur généralise bien sur les données des 4 participants appartenant au groupe TUN=1/0 n'ayant pas été utilisées pour l'apprentissage. Ceci confirme la pertinence de l'approche pour une application dans le domaine des facteurs humains.

ENGLISH VERSION

Our goal is to find ways to detect the operator attentional state in real time so that dynamic adaptations of the human-machine system can avoid fatal accidents. In some situations, attentional tunneling is likely to appear and leads to excessive focus and poor decision making. The allocation of visual attention is a key factor when operating complex systems under time pressure. As presented in the previous chapters, eye-tracking provides interesting metrics for detecting attentional impairments as attentional tunnelling.

Metrics related to attentional tunnelling presented in the previous chapter were implemented to objectively measure participants behaviour. We combine an AOI-based approach and heart rate as an index of attentional tunnelling. Using a classical statistical approach, we show that these metrics are efficient to discriminate between participants who faced attentional tunnelling and the others.

This result paved the way to the implementation of machine learning techniques to automatically infer attentional tunnelling. We used Adaptive-Network-based Fuzzy Inference System (ANFIS) which is a technique particularly suited to process noisy and continuous data. Moreover ANFIS provides fuzzy rules that can easily be interpreted by a human factor expert. The whole training of the ANFIS is presented and the test of the classifier on new data demonstrates that the fuzzy inferred rules can be generalized. This work was published following in IEEE Transactions on Human-Machine Systems journal (Regis et al., 2014).

5.1 OBJECTIVES OF THE STUDY

5.1.1 *Metrics of Attentional Tunneling*

Considering Wickens' definition of attentional tunneling (Wickens, 2005), a straightforward way to identify this phenomenon is noting when operators omit unexpected events (e.g., they do not react to alarms) and persevere in their current action pattern. Such an expert approach requires analysis of the operators' behaviors to infer their attentional state (e.g., actions on the user interface reaction time). A complementary approach is to derive attentional tunneling from the measurement of physiological signals and ocular activity. Indeed, this attentional impairment is associated with psychological stress (Bahrick et al., 1952, Weltman et al., 1966, Easterbrook, 1991). Several authors have demonstrated that attentional tunneling results in fewer scanned areas of interest (AOI) on the user interface (Wickens, 2005), a decreased saccadic activity (Tsai et al., 2007), long eye fixations (Cowen et al., 2002), and the absence of ocular fixations on relevant cues (Thomas and Wickens, 2004).

5.1.2 *Formal Inference Techniques*

The efficiency of attention tunneling identification depends not only on the selection of accurate metrics but also on the type of classification technique. Moreover, classifying attentional tunneling states on the basis of physiological and behavioral metrics requires some flexibility. Physiological metrics are generally continuous and noisy and there are no mathematical models of the links between physiology and attention. Only expert models are generally used to link physiology and attention. Fuzzy logic is an inference technique that is well suited for continuous and noisy inputs (Cox, 1992). Following a similar approach proposed by Mandryk and Atkins (2007), Pizziol et al. (2011) developed a mathematical model using fuzzy rules to link psychophysiological inputs (e.g., the heart rate (HR) used as a psychological stress indicator) and an attentional output (e.g., "the level" of attentional tunneling). Although this study provided promising results showing the interest of fuzzy logic for such modelling, there was a limitation to this approach. Indeed, the fuzzy rules that link these inputs and outputs had to be set a priori from expertise. A consistent method to avoid such drawbacks is to use automated machine-learning techniques (Hastie et al., 2005), as they allow mathematical links between inputs and outputs to be established from a statistically sound point of view, rather than relying on experts. The machine-learning literature provides a wide range of methods and algorithms to learn efficient classifiers (e.g., neural networks Wilson and Russell (1999), support vector machines (SVMs) Bergasa et al. (2006), hidden Markov models Codispoti et al. (2001)) from which one can find the appropriate method for the specific application. Neuro-fuzzy learning Jang et al. (2009) is well suited in this particular case as in addition to its learning ability, the method retains the advantages of fuzzy logic. Furthermore, under certain conditions (Riid and Rüstern, 2000), the fuzzy rules underlying the behavior of the generated classifier can be translated into natural language (Zadeh, 1996), thus allowing an easier interpretation by researchers.

5.1.3 *Present Study*

The main objective of this research is to provide a formal method using machine-learning inference techniques to detect attentional tunneling from the analysis of psycho-physiological and oculomotor responses that are collected during an experiment Dehais, Causse and Tremblay (2011a). This experiment was designed to provoke attentional tunneling. Participants were asked to perform a task of manually controlling a robot while performing target identification. During the task, a low battery failure was triggered, forcing a safety procedure that required the robot to return autonomously to the base. At

this stage, the participants were supposed to stop operating the robot and allow it to return back to the base. The participants were separated in two groups: the control group, whose members did not receive any cognitive countermeasures to help them notice the battery failure, and the countermeasure group, whose members were assisted with a cognitive countermeasure. In this experiment, the battery failure was used as a probe to determine whether the participants faced attentional tunneling or not, in accordance to Wickens' definition Wickens (2005). This approach allowed labeling of attentional tunneling periods for supervised machine-learning purposes. Three attentional tunneling metrics were employed in this study: the HR, the number ofAOI (NBAOI) glanced at on the user interface, and the switching rate (SWR). The literature suggests that HR increases with psychological stress (Pomeranz et al., 1985, Warner and Strowman, 1995, Parati et al., 1995, Causse, 2010, Causse et al., 2011, Dehais, Causse and Tremblay, 2011a, Causse et al., 2013) and NBAOI and SWR decrease with attentional tunneling Tsai et al. (2007), Cowen et al. (2002). The calculation of the attentional tunneling periods and the changes in the three associated metrics are presented in Section III. These three metrics were used to train an adaptive-networkbased fuzzy inference system (ANFIS) to infer attentional tunneling according to the labeling that is used in the experiment. The ANFIS was chosen as it combines the expressive power of fuzzy representations with the adaptability of neural networks for more accurate predictions. The performance of this inference system was assessed in terms of learning and robustness. A support vector machine was implemented to compare classification performances of these two algorithms. The methodology to train the ANFIS and the performance analysis on the robotic experiment is described in Section IV.

5.2 EXPERIMENTAL METHODS

5.2.1 *Material*

The experimental setup consisted of a robot that is equipped with different sensors and a ground station that is used to control the robot. The robot can be operated in the "manual" or "supervised" mode. In the manual mode, the robot was controlled by the operator with a joystick. In the supervised mode, the robot performed waypoint navigation autonomously, but any action of the operator with the joystick allowed him/her to take over until the joystick was released. The ground station (see Figure 22) used a 24-in display that provided information for control and supervision of the robot. The operator could not see the robot and only gathered information regarding the scenario through the display interface.

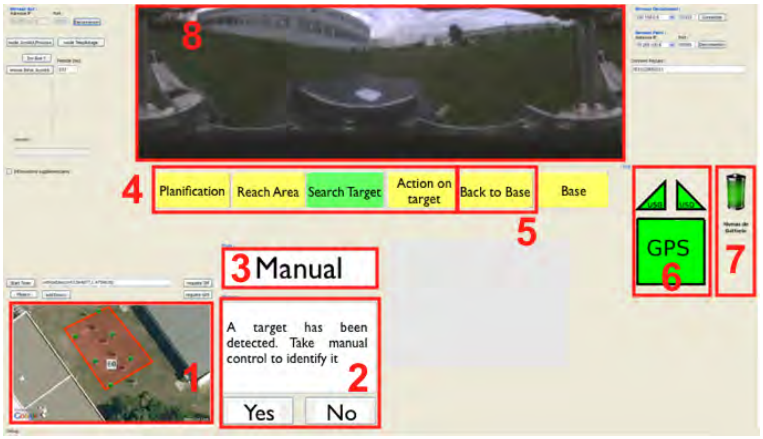


Figure 22: Ground Station Interface: 1) "tactical map", 2) "interactive panel", 3) "mode annunciator", 4) "synoptic", 5) "back to base", 6) "GPS and ultrasound status", 7) "battery status", 8) "panoramic video".

5.2.2 *Experimetal Scenario*

The experimental scenario was designed to induce attentional tunneling. The scenario consisted of a target localization and identification task. The target was a black-metal panel with red stripes and two messages written in white: "OK," (front side) "KO" (backside). The mission was 4 min long and composed of four main segments: S1-"Reach the area," S2-"Scan for target," S3-"Identify target," and S4-"Battery Failure." At the beginning of the mission, the operator navigated the robot in the supervised mode to reach the search area (S1). Upon arrival, the robot began detecting the target (S2). When the robot was within the vicinity of the target, the operator was notified to control the robot in the manual mode for identification and differentiation of both messages (S3). The use of a panoramic video [35] and the introduction of a 1-s lag in the control loop (Beigbeder et al., 2004) increased the task difficulty and favored excessive focus. While the operator was involved in the identification task (S3), a "low-battery event" was sent by the experimenter (S4). This event triggered a safety procedure that forced the robot to automatically return to base in the supervised mode unless the operator overrode this procedure with a command from the joystick.

As this failure happened at a crucial moment in the mission when the operator was particularly committed to handling the robot near the target, it was expected that the operator would not notice the alerts on the interface warning of the "low-battery" event and would persist in achieving the target detection task. Thus, the operator would experience attentional tunneling.

5.2.3 *Participants*

Twenty-three participants (all males but four females, mean age = 29.52, SD = 9.14), all French defense staff from Institut Sup´erieur de l'A´eronautique and de l'Espace and The French Aerospace Lab were recruited by local flyers and emails. Informed consent was received at the start of the experiment. Participants were randomly assigned to two independent groups.

1. The control group consisted of 12 participants (all males but two females, mean age=28.25, SD=6.64). No countermeasure was provided to help them detect the battery failure.

2. The countermeasure group consisted of 11 participants (all males but two females, mean age = 30.90, SD = 11.45). A countermeasure was provided to help them detect the battery failure.

5.2.4 *Procedure*

The participants sat 1 m from the user interface in a closed room with no visual contact with the outdoor playground where the robot moved.AProComp Infinity system (Thought Technology) electrocardiogram was applied to the participants' chests using Uni-Gel to enhance the quality of the heart beat signal. A Pertech head-mounted eye-tracker was placed on their heads to observe their oculomotor behaviors. Participants rested for 3 min to establish a physiological baseline and performed a 13-point eye-tracking exercise to calibrate both sensors. A briefing was provided on the mission, the user interface, and the two guidance modes. Participants were trained for 20 min on controlling the robot through the panoramic video screen with the two guidance modes. They were told that four incidents (low-battery event, communication breakdown, GPS loss, and ultrasound sensors loss) might occur during the mission. However, only the low-battery event occurred during the experiment. The associated procedures and the expected robot behaviors were explained for each of these incidents. During the low-battery event, participants were to "release immediately the joystick and let the robot go back to base"; during the communication breakdown and the GPS failure, they were to "wait for the communication or the GPS signal to return"; and during the ultrasound sensor failure, they were to "use the joystick to avoid obstacles." The participants were also trained on how to diagnose these four issues from the user interface. The low-battery event caused three main changes in the user interface (see Figure 23): the battery icon turns to orange with an associated "low battery" message (see Figure 22, area 7), themode changes frommanual to supervised and flashes twice (see Figure 22, area 3), and the segment status becomes back to base (see Figure 22, areas 4 and 5). There was only

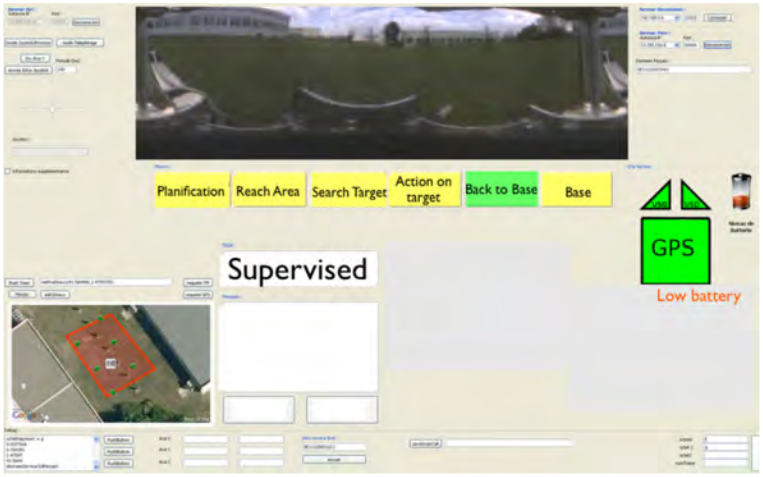


Figure 23: Interface as the battery failure is triggered

one change each for the other scenarios - communication breakdown: "the panoramic video is frozen"; GPS loss: "the GPS icon becomes red and the mode changes to manual"; ultrasound sensor failure: "the ultrasound icons turn to red." After the training session, there was a short test to verify participants' understanding of the instructions and procedures. The experimental scenario that involved the battery failure was initiated and only occurred in one trial per participant.

5.2.5 *Failure and Cognitive Countermeasure*

As described in Section II-B, the low-battery event triggered an automatic procedure that forced the robot to return to the base by the shortest route. The participants were informed of the occurrence of this event through the user interface as described in Section II-D. A cognitive countermeasure was designed to help the participants of the countermeasure group to deal with attentional tunneling. It was hypothesized that the operators would be excessively focused on the panoramic video window for target identification (see Figure 22, area 8). This window was removed for 1 s, replaced by the explanation of the robot behavior for 3 s, and reappeared with the explanation superimposed for three more seconds. Finally, the interface returned to its nominal state. The robot did not move while the cognitive countermeasure was sent to the operator.

5.2.6 *Metrics Computation*

1. Heart Rate: The electrocardiograph was sampled at 2048 Hz. The BioGraph Infiniti software was used to export the HR computed from the interbeat-interval (R–R interval) every second of the experiment. No complementary spectral analysis meth-

ods were used because of the length of the time window required to process this kind of analysis (at least 4 to 5 min (Parati et al., 1995)): only a moving average filter of 5 s was applied to smooth the HR signal. Because of a commonly observed difference in HR baseline values among participants, HR values were standardized: For each participant, the mean HR of the resting period was subtracted from the HR data (Bonner and Wilson, 2002). Therefore, the relevant metrics was the difference between the current HR and the baseline, expressed in beats per minute.

2. Number of Areas of Interest: The EyeTechLab eye tracking software provided timestamped data of the cartesiancoordinates of the participants' eye gazes on the visual scene at a 25-Hz sampling rate (40 ms between samples). Eight AOI were defined on the user interface as presented in Figure 22, identified by red rectangles. Each eye position was labeled in accordance to the most relevant AOI; if the eye position was not in an AOI perimeter it was labeled as "AOI number 0." The NBAOI was computed every second of the experiment. Each value of NBAOI was the number of AOIs used during the last 20 s.

3. Switching Rate: Similarly to NBAOI, SWR relied on the eye tracking data. SWR was also computed every second, and corresponded to the number of gaze transitions from one AOI to another during the last 10.5 s, expressed in number of transitions per minute. Pizziol et al. (2011) provide a more precise definition of NBAOI and SWR.

4. Data Collection: The raw measurements that constitute the three metrics were collected with no online processing. The metrics were computed offline and synchronized. For each participant, the final data were stored in a timestamped log as a sequence of triplets (HR, NBAOI, and SWR), with each triplet covering 1 s of experiment. Two examples of metrics recordings can be found on Figure 30 and Figure 31.

5.3 EXPERIMENTAL RESULTS

5.3.1 *Behavioural Results and Expert Labeling of the Periods of Attentional Tunneling (TUN)*

The results of the control group revealed that eight participants out of 12 (66.67%) experienced attentional tunneling: They persisted in examining the target instead of letting the robot go back to base. Although they felt surprised by the behavior of the robot, these participants all declared that they neither noticed the low-battery event

nor the other changes on the user interface. The other four participants reported that they had noticed the failure and had decided to let the robot go back to base. These subjective results were consistent with the oculomotor measurement that revealed that these four participants glanced at the battery icon prior to releasing the joystick. These participants achieved the appropriate situation awareness with no help and conducted the mission successfully. In contrast, all 11 participants from the countermeasure group noticed the battery failure and understood the behavior of the robot. The eye-tracking analysis showed that all of the participants who gazed at the alarm released the joystick to let the robot go back to base. In this group, 10 out of 11 participants made the decision to stop the mission and let the robot go back to base in the supervised mode. Only one deliberately persisted in identifying the target for 50 s until the battery failed and therefore did not achieve the mission. That participant believed that the remaining power was enough to finish the mission despite the occurrence of the battery failure. His data were not used in this study.

There is no straightforward method to measure the level of attentional tunneling. However, thanks to an expert approach as described in Section I-A, small periods could be labeled with an attentional tunneling level (TUN) of 1 or 0 during the last segment of the experiment (S4). Indeed, according to Wickens' definition, the detection of the unexpected battery failure event was used as an attentional tunneling probe [6] which led to two types of labels:

1. TUN = 1: Eight samples were labeled with TUN = 1 from the failure until the end of the experiment. They correspond to the eight participants from the control group who did not glance at the battery failure icon and persisted in examining the target instead of letting the robot go back to base. These samples were used to train the ANFIS.

2. TUN = 0: Ten samples were labeled with TUN = 0 from 20 s after the failure until the end of the experiment. They correspond to the ten participants from the countermeasure group, who noticed the failure immediately and followed the associated procedure (let the robot go back to base in supervised mode). Discarding the first 20 s nullifies noise because of the cardiac response latency [38] after the failure. These samples were used to train the ANFIS.

3. TUN = 1/0: Four participants were labeled with TUN = 1 until they glanced at any of the following: the mode annunciator (AOI3), the synopsis (AOI4), the "back to base" sign (AOI5), and the battery status (AOI7). They were labeled with TUN = 0 afterward until the end of the experiment. They correspond to the four participants from the control group who reported that they had noticed the failure and had decided to let the robot go

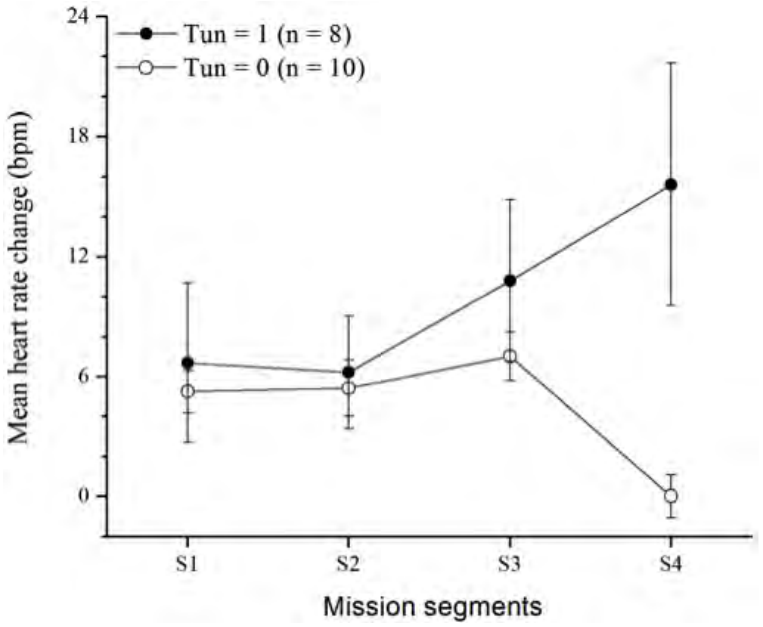


Figure 24: Mean heart rate changes according to the four mission segments for each category of participants. Error bars represent the standard error of the mean.

back to base. These data were used neither for the inferential analysis nor for the training of the ANFIS but only to test the model.

5.3.2 *Inferential Analysis of the Input Raw Data*

The approach thus far has consisted of defining the three different physiological and oculomotor metrics (HR, NBAOI, and SWR) that characterize the psycho-physiological evolution of each participant along the experiment. Parametric repeated measures ANOVA and Fisher's least significant difference test were used to examine the effects of the mission segment type on the three metrics. A categorical explanatory variable was used in the analysis to check for differences between the two categories of participants depicted previously, TUN = 0 (n = 10) and TUN = 1 (n = 8). The four participants from the TUN = 1/0 group were not included in the analysis.All testswere conducted at $\alpha = 0.05$.

1. Heart rate (see Figure 24): The two-way repeated-measures ANOVA showed a significant group×segment type interaction, $F(3, 48) = 13.60, p < 0.001, ETAcarrep = 0.46onHR$. Paired comparisons indicated that the mean HR change differed between the TUN = 0 and the TUN = 1 groups during S4 (p = 0.001), which is the segment containing the battery failure. HR significantly declined between S3 and S4 in the TUN = 0 group (p < 0.001); on

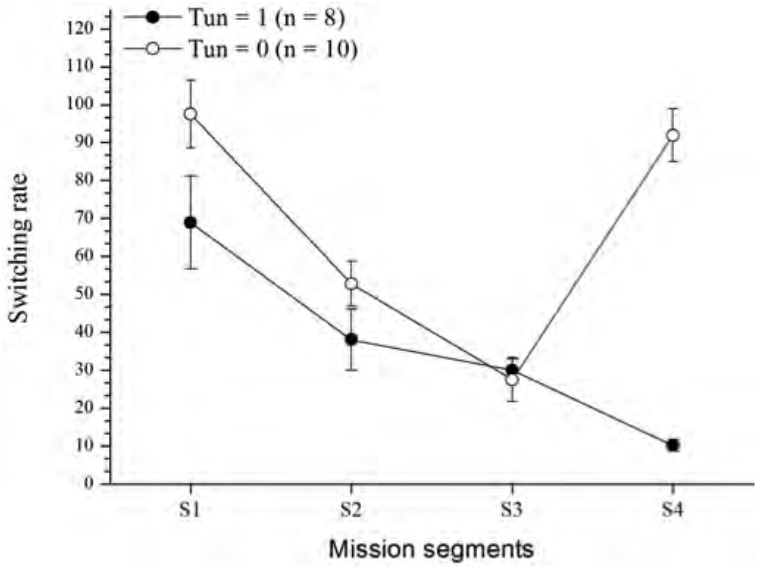


Figure 25: Mean AOI switching rate according to the four segments for each category of participants. Error bars represent the standard error of the mean.

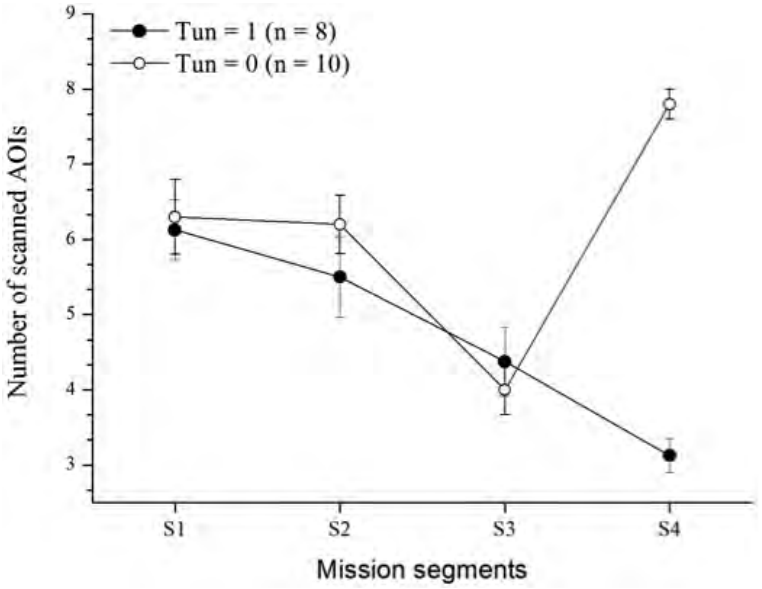


Figure 26: Mean number of scanned AOI according to the four segments for each category of participants. Error bars represent the standard error of the mean.

the contrary, it significantly increased between S3 and S4 in the TUN = 1 group ($p = 0.018$).

2. Number of scanned AOIs (see Figure 26): There was a significant group × segment type interaction, $F(3, 48) = 21.35$, $p < 0.001$, ETA 2 p = 0.57 on NBAOI. Paired comparisons showed that the number of scanned AOIs differed between the TUN = 0 and the TUN = 1 groups during S4 ($p < 0.001$). Paired comparisons revealed that the NBAOI significantly increased from S3 to S4 in theTUN=0 group ($p < 0.001$), whereasthis value dropped in the TUN = 1 group ($p = 0.021$).

3. Switching rate (see Figure 25): The same analysis performed on the gaze SWR (gaze transitions from AOI to AOI per minute) also revealed a group × segment type interaction, $F(3, 48)=16.98$,p < 0.001, ETA 2 p =0.51) on SWR. During S4, the transition rate differed between the TUN = 0 and the TUN=1 groups ($p < 0.001$). In the latter segment, the mean transition rate increased drastically in the TUN = 0 group ($p < 0.001$), whereas it decreased in the TUN = 1 group ($p = 0.038$). These results are consistent with the previous literature on attentional tunneling. Participants from the TUN = 1 group were highly focused on the demanding identification task during S4 as revealed by the decrease of the NBAOI and SWR. Their inability to successfully perform this critical task generated stress (Folkman, 1984), associated with the higher HR during S4 (Pomeranz et al., 1985, Warner and Strowman, 1995, Parati et al., 1995, Causse, 2010, Causse et al., 2011, Dehais, Causse and Tremblay, 2011a, Causse et al., 2013). Conversely, participants from the TUN = 0 group did not faced such stress. Their HR was relatively lower during S4. One has to consider that these participants noticed the alarm and just let the robot go back to base autonomously as stated by the procedure. Consistently with (Wickens, 2005, Tsai et al., 2007, Cowen et al., 2002), NBAOI decreased on S4 for the TUN = 1 group (participants facing attentional tunneling consulted less instruments) and raised for the TUN = 0 group. In addition, the SWR decreased for the TUN = 1 group on S4 (participants facing attentional tunneling had a reduced saccadic activity) as opposed to the TUN = 0 group.

5.4 ATTENTIONAL TUNNELING DETECTION

This section presents the machine learning method that was most appropriate for this application, ANFIS, and focuses on its implementation and tuning. Machine learning aims at automatically generalizing abstract concepts from numerical experimental data. In particular, assigning discrete labels to different measurements, given previous ex-

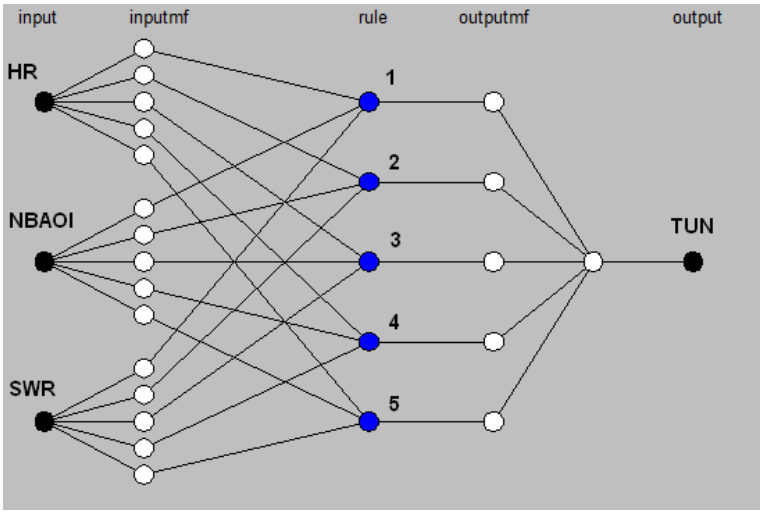


Figure 27: Example of an adaptive network-based Sugeno-type fuzzy inference system

amples of correctly labeled measurements, is known as classification within the supervised statistical learning literature Hastie et al. (2005). Assessing whether a person is currently facing attentional tunneling, given the current values of the metrics and previous examples, is hence a classification task. The classification task at hand consists in finding a mapping from the triplets x = (HR, NBAOI, and SWR) to a label TUN in 0; 1.

5.4.1 *Adaptive-Network-Based Fuzzy Inference System to Classify Attentional States*

An ANFIS classifier is a five-layered neural network using fuzzy membership functions for the activation neurons of the input layer [hence the name of "fuzzy inference system (FIS)"]. More specifically, as shown in Figure 27, in the first-order Sugeno fuzzy model (Sugeno and Kang, 1988) that was employed, the following five layers are used.

1. The first layer (see "inputmf" layer in Figure 27) is a fuzzyfication level where each neuron corresponds to a membership function; it takes the metrics as inputs and outputs the values of the membership functions. The training phase tunes the parameters of these membership functions (e.g., their center position).

2. The second layer introduces nonlinear logical dependences between fuzzified metrics by combining the outputs of the first layer. This combination is done using "and" rules that multiply their inputs together.

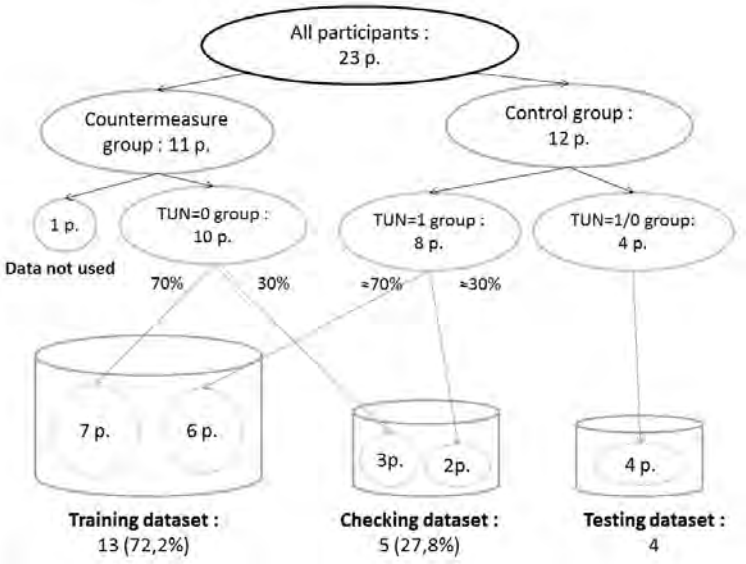


Figure 28: Data distribution between training, checking, and testing datasets.

3. The third layer takes all output values from the second layer and normalizes them so that they sum to one. In Figure 27, the second and third layers have been merged into the "rule" layer for clarity. Note that these two layers have no tunable parameters.

4. Then, the fourth layer (see "outputmf" layer in Figure 27) is a recombination layer between normalized rules that multiplies the output of the third layer by a first-order polynomial of the metrics (hence the name "first-order Sugeno model"). Its neuron parameters are the polynomial coefficients. Note that, for the sake of clarity, the direct connections from the input layer to "outputmf" are not drawn on Figure 27.

5. Finally, the fifth layer holds a single neuron in which a weighted average of all outputs from the fourth layer results in the ANFIS continuous output value.

The MATLAB fuzzy logic toolbox was used to perform the calculations. As recommended by Jang and Sun [25], the training algorithm used is a hybrid combination of least-squares fitting (for the outputmf parameters) and back propagation gradient descent (for the inputmf parameters). Contrary to a zeroorder Sugeno fuzzy model, the output value in a first-order Sugeno is input-dependent because of the first-order polynomial of the metrics. This feature allows for more flexibility in modeling the relationships between the inputs and the output but threatens the transparency of the model (i.e., its interpretability) (Riid et al., 2001, De Oliveira, 1999).

5.4.2 *Training Data*

ANFIS requires a reference dataset of pairs (x, TUN) where x is the triplet of psycho-physiological parameters (HR, NBAOI, and SWR) and TUN is the level of attentional tunneling. The only appropriate samples are the labeled samples from the segment 4 (cf. Section III-A). Only the samples from the eight participants from the TUN = 1 group and from the ten participants from the TUN = 0 group were used to train the ANFIS and constituted the reference dataset. Cross validation was performed to control the ANFIS training in order to improve its generalization. The reference dataset had to be split in two smaller datasets: the training dataset and the checking dataset. A 70/30% balance was achieved: six out of the eight "TUN = 1" participants and seven out of the ten "TUN = 0" participants were randomly selected and their labeled samples constituted the training dataset (representing 72.2% of the 18 participants). The data from the other two TUN = 1 participants and the three TUN = 0 participants formed the checking dataset (representing 27.8% of the 18 participants). The data from the four remaining participants were used later on to test the ANFIS classifier on new data and evaluate its robustness after the training. They constituted the testing dataset. A summary of the data distribution is presented in Figure 28.

5.4.3 *Parameters of the Fuzzy Inference System*

The first step before training the ANFIS is to choose the structure of the FIS itself. This structure depends on the number of inputs and on the number of membership functions for each input. The classical approach at this stage is to use grid partitioning to define the number of membership functions for each input. The subtractive clustering method was used instead in order to let the algorithm discover a statistically sound number of clusters in the training data without biasing it with expert knowledge. The other advantage of this method is that it considerably reduces the number of rules that have to be trained by the adaptive neural network, which results in a faster computation and an easier interpretation of the rules after training. Yager and Filev explain in greater detail the mountain clustering method from which subtractive clustering is derived Yager and Filev (1994). According to this method,1 five clusters have been elicited, covering all the samples in the input three-dimensional (3-D) space. Each cluster is a 3-D set, decomposed in one membership function per input dimension. The resulting FIS is therefore composed of five rules corresponding to the five rules to be trained as presented in Figure 27.

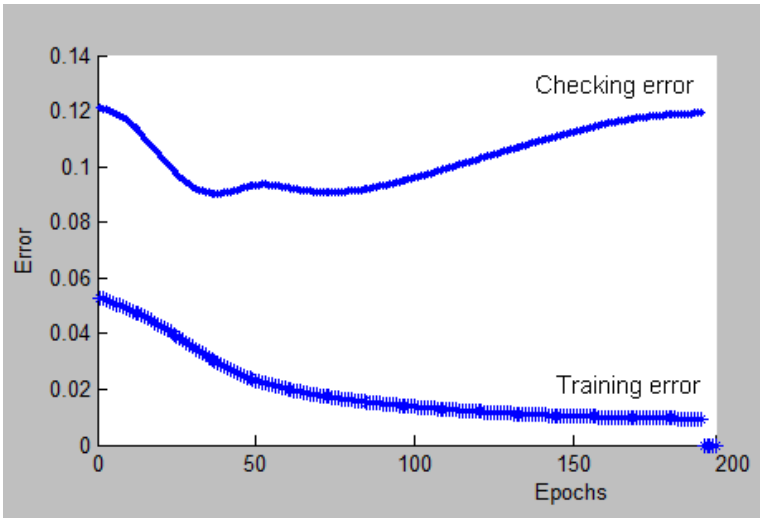


Figure 29: Error evolution during the training process. The error on the training dataset constantly decreases (bottom curve), while the error on the checking dataset begins to increase after the 38th epoch.

5.4.4 *Training of the Adaptive-Network-Based Fuzzy Inference System*

The training of the ANFIS is possible once the FIS has been set. The tunable parameters of the ANFIS are updated during a repetition of training epochs. Each training epoch is therefore a local optimization of the ANFIS ability to classify the training dataset properly. However, neural networks such as ANFIS are prone to overfitting (because of excessive training epochs, the ANFIS becomes specific in modeling the training dataset, but performs poorly on new data). To avoid this, cross validation was performed: After each training epoch that is based on the training dataset, the ANFIS classified the checking dataset and the classification error was evaluated (checking error). Overfitting appeared after 38 training epochs, as the trained ANFIS did not improve in classifying the checking data despite new training epochs (see Figure 29). Training is therefore stopped at the 38th epoch. The resulting training error is 2.9% corresponding to the minimum of the checking error (8.9%).

5.4.5 *Performance on the Training Data*

The performance of the ANFIS was evaluated on labeled samples from the segment S4 only, where the ANFIS TUN level prediction can be compared with the expert TUN level. There is a slight dissymmetry in the fitting between training data and the ANFIS prediction. If one separates (TUN = 0) and (TUN = 1) participants within the training set, then the false positive (type II error) rate is 4% and the false negative (type I error) rate is 0%. Two examples of the ANFIS output

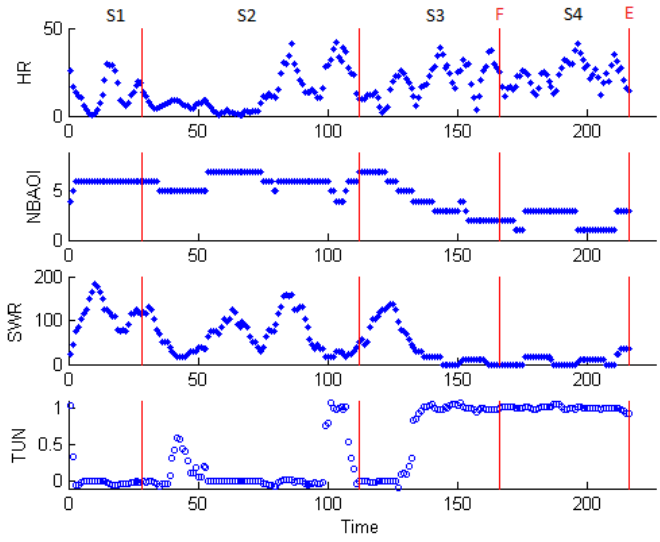


Figure 30: Evolution of the three metrics for the participant DUPNI (from the TUN = 1 group) during the whole experiment and associated TUN level inferred by the trained ANFIS. F: battery failure time. E: end of the mission.

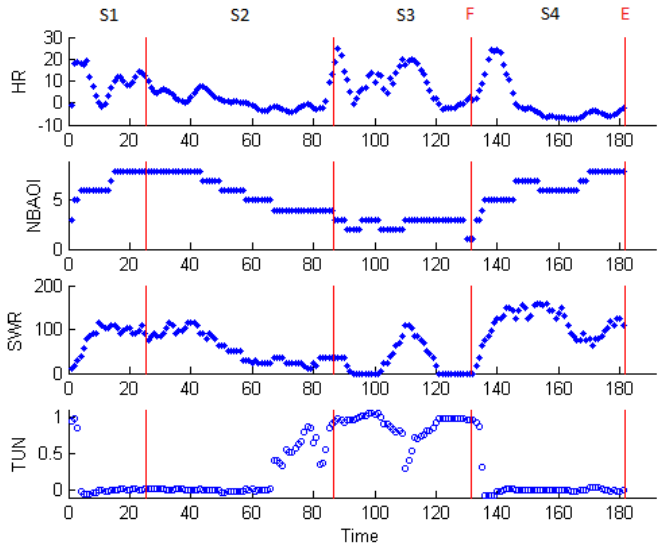


Figure 31: Evolution of the three metrics for the participant JACRA (from the TUN = 0 group) during the whole experiment and associated TUN level inferred by the trained ANFIS. F: battery failure time. E: end of the mission.

Kernel Function Type	Checking Error
Linear	1.9%
Quadratic	2.6%
Polynomial	3.8%
Radial Basis Function	2.3%

Table 3: SVM cross validation error with different kernel function types

on the training data can be found in Figs. 9 and 10. The TUN level was computed with the trained ANFIS for the whole mission. In Fig. 9, the HR rose around the end of S2 and kept a high variability across S3 and S4. NBAOI remained between 5 and 7 before S3, progressively dropped during S3 and remained around 2 till the end of the experiment. SWR also dropped during S3 to reach values close to 0 until the end of the experiment. The TUN level increased during S3 and remained close to 1 until the end of the experiment. These evolutions supported the expert hypothesis: Participant DUPNI experienced attentional tunneling during the all of S4. In Fig. 10, HR rose abruptly at the beginning of S3 and S4. NBAOI slowly decreased from values around 8 during S2 to reach its minimum values around 2 during S3 and rose back to values around 7 during S4. SWR follows a similar evolution as NBAOI. These evolutions supported the expert hypothesis: Participant JACRA did not face attentional tunneling during the whole segment 4.

5.4.6 *Support Vector Machine Comparison*

Using the same training and checking datasets, the same classification task was performed with a SVM trained with the Bioinformatics toolbox in MATLAB. Here, the SVM is used as the reference method to compare the results obtained with the ANFIS. For more details about the SVM, see Cortes and Vapnik (1995), Vapnik (1999, 2000), Smola and Schölkopf (2004). The results of the SVM cross validation, with different kernel function types, are presented in Table 3.

The linear kernel performs best in this experiment; therefore, this kernel was employed for the SVM. Comparing and analyzing the checking error differences between the different kernels is beyond the scope of this paper.

In order to compare the predictions of both methods in terms of binary outputs (0 or 1, instead of continuous values), the continuous ANFIS outputwas rounded to the nearest integer and only the sign of the SVM output was considered. The rounded predictions of the ANFIS have been compared with the expert TUN values on the checking dataset and an error of 1.1% was found. It is a better score than the 1.9% obtained with the linear kernel SVM (see Table I).

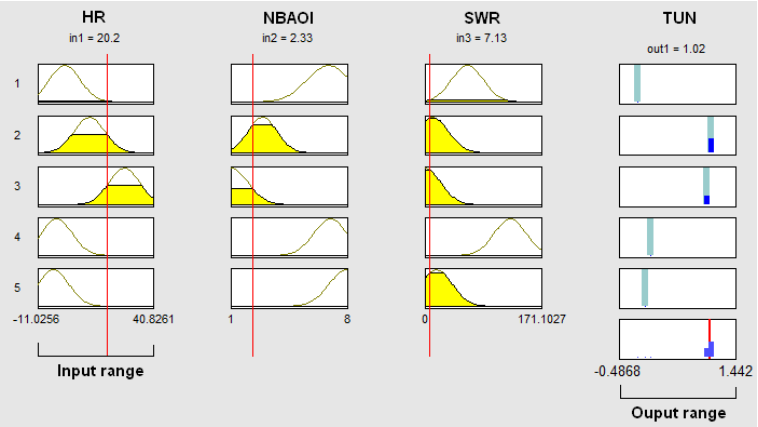


Figure 32: Clusters and rules representation of the trained ANFIS. the first three columns represent the three inputs. Each column is composed of five lines corresponding to the five clusters elicite by the substractive clustering algorithm. the gaussian curves represent the associated membership functions. The yellow levels (between 0 and 1) are the membership function activations depending on the three input values (vertical red lines). The activations of the output rules are computed lini by line, depensing on the activation of the three input mebership functions (a rule is fully activated when the three membership functions are fully activated). the contribution of each output rule is then defuzzified in the right column. fianlly, the TUN output is computed via a weighted sum of the five output contributions (vertical red line in the bottom right rectangle).

Furthermore, the prediction discrepancy between the two classifiers was studied in order to verify whether there was agreement when one was misclassified an example. It appeared that all elements misclassified by the ANFIS were also misclassified by the SVM predictor. Therefore, both methods exhibit consistent behavior. Consequently, for the 1.1set misclassified by the ANFIS, one can conjecture that the checking examples are either outliers or exhibit values that are too different from the training set to be classified accurately.

5.4.7 *Adaptive-Network-Based Fuzzy Inference System Analysis*

5.4.7.1 *Adaptive-Network-Based Fuzzy Inference System Rules Interpretation:*

The clusters determined by the subtractive clustering method and a representation of the trained fuzzy rules can be found in Figure 32.

As mentioned in Section IV-A, contrary to a zero-order Sugeno model, the transparency is not granted with a first-order Sugeno. It appears that the first-order features in this model are negligible when compared with their constant counterparts (i.e., the first-order poly-

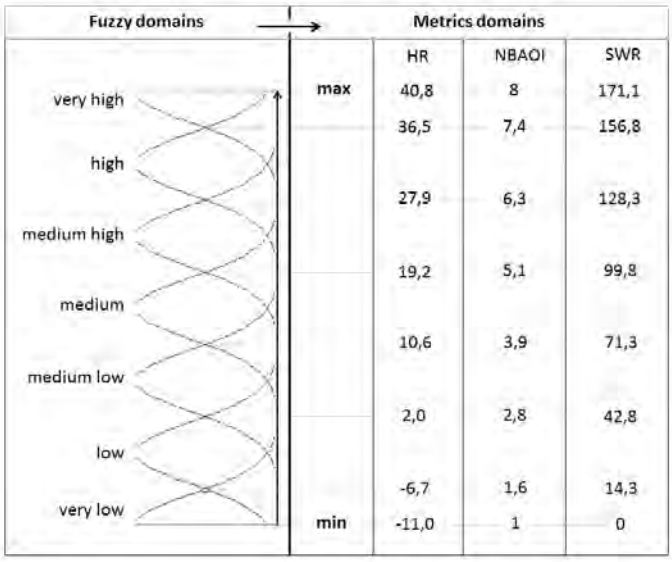


Figure 33: Fuzzy domain definition used to translate the ANFIS rules in natural language and associated thresholds in the metrics domains.

nomial coefficients in the fourth layer are negligible when compared with the zero-order coefficients). Therefore, this model behaves almost like a zero-order Sugeno and preserves its interpretability from a linguistic point of view.

In order to translate the five fuzzy rules in the natural language, seven fuzzy domains were defined from "very low" to "very high" (see **??**). Among the five rules that are trained by the ANFIS, three of them "pull" the TUN output to 0 when they are activated (see rules 1, 4, and 5, Figure 32). These rules are activated when:

1. (HR is low) and (NBAOI is high) and (SWR is medium low);
2. (HR is low) and (NBAOI is high) and (SWR is medium high);
3. (HR is low) and (NBAOI is very high) and (SWR is low).

The remaining two (rules 2 and 3) that "pull" the output to one are activated when:

1. (HR is medium) and (NBAOI is low) and (SWR is very low);
2. (HR is high) and (NBAOI is very low) and (SWR is very low).

5.4.7.2 *Metrics Importance in theAdaptive-Network-Based Fuzzy Inference System Performance:*

In order to evaluate the role of the three metrics used as inputs in the diagnosis of attentional tunneling, the performance of the previously trained three-input classifier was compared with three classifiers trained with a combination of only two inputs out of three. The

Set of inputs	Epochs	Training error	Checking error
All three inputs	38	2.9%	8.9%
Without HR	59	6.5%	11.8%
Without SWR	95	7.8%	15.2%
Without NBAOI	>100	9.6%	21.2%

Table 4: Cross validation error comparison between the two-input and the three input classifiers

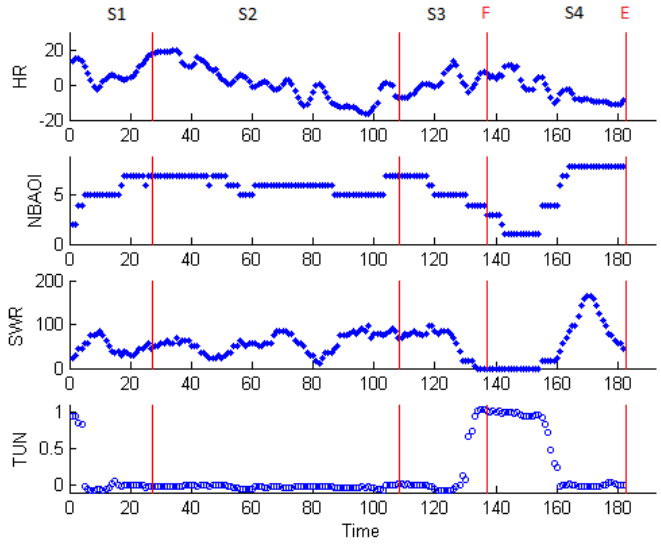


Figure 34: Evolution of the three metrics for the participant HOSAL from the TUN=1/0 group and associated TUN level computed by the trained ANFIS. F: battery failure time, E: end of the mission.

same training method was used to train these new classifiers. The number of epochs at which the checking error increased varied from one classifier to another. The results presented in Table 4 reveal that the three metrics contributed to the attentional tunneling classification as the checking error is minimal when using them together as inputs. NBAOI appeared to be the most important metric among the three as the checking error raises to 21.2% when training the ANFIS without this metric. SWRalso played a more important role in predicting attentional tunneling than HR, which appeared to only slightly improve the diagnosis (checking error is 11.8% without HR compared with 8.9% when added as the third input).

Participant	Expert	ANFIS
Participant ID	TUN level at failure T when Gaze at alarm cues T when release joystick	TUN level at failure Time when TUN=0
PENSE	1 18s (gaze alarms) 45s (release joystick)	1.026 17s
NIVAL	1 20s (gaze alarms) 30s (release joystick)	1.019 27s
HOSAL	1 20s (gaze alarms) 22s (release joystick)	1.03 22s
DASJE	0 0s (gaze alarms) 16s (release joystick)	-0,02 0s

Table 5: Comparison between expected TUN level and ANFIS classification. For example, participant NIVAL was considered as facing attentional tunneling (TUN=1) at the moment of the failure and remained in that state 20s. before he noticed the alamr (expert point of view). The ANFIS classifier output was at TUN=1.019 at the moment of the failure and abruptly dropped to values close to 0 after 27s.

5.4.8 *Adaptive-Network-Based Fuzzy Inference System Testing: Classification Over New Data*

In order to check the robustness of the ANFIS classifier, it was tested on the four participants from the TUN = 1/0 group who detected the failure before the end of the experiment, and whose data were used neither for training nor for checking. Consistently with the expert approach used to label the samples (see Section III-A), we first determined from the data:

1. the level of attentional tunneling at the moment of the failure; and
2. the time when the participants switched from attentional tunneling (TUN = 1) to a nominal behavior (TUN = 0).

The comparison between the latter expert results and the outputs of the ANFIS classifier appears to be consistent as presented in Table 5. An example of the inference of the level of attentional tunneling of one of these four participants can be found in Figure 34.

5.5 CONCLUSION OF THE STUDY

After validating that our metrics of choice were sensitive to attentional tunneling, we proposed to implement the ANFIS method for automating the detection of attentional tunneling. This formal approach was tested a posteriori on data collected during an experiment conducted in the domain of human operator–robot interactions. The training of the algorithm was performed using samples from the last segment of the mission when a failure occurred. The absence of perception of the visual alerts associated with the failure and the persistence of the participants in their initial goal were used as objective indicators of attentional tunneling. The inputs of the ANFIS consisted of a set of three physiological and oculomotor parameters that are known to be associated with attentional tunneling in the human factor literature.

For the machine-learning purpose, the first objective was to ensure that it was possible to collect data from participants who faced attentional tunneling versus participants who did not. The experiment was successful in generating these two types of behaviors. Indeed, eight participants faced attentional tunneling from the battery failure until the end of the experiment, whereas ten did not during the same segment. These behavioral results were consistent with the psychophysiological and oculomotor measurements. Indeed, during S4 (segment after failure), the group who faced attentional tunneling (TUN = 1) exhibited a higher HR, a lower saccadic activity, and fewer scanned AOI than the group who noticed the alarm (TUN = 0). Furthermore, the inferential analysis on the three metrics showed significant segment

type×group interaction that validated the choice of the three metrics to derive attentional tunneling. Indeed, the group who faced attentional tunneling (TUN = 1) had a significantly higher HR, a lower saccadic activity, and fewer scanned AOI on S_4 than on S_3 even though the task they performed over these two segments was similar. This result highlighted a change in the metrics that was not accounted for a change in the task but that was consistent with the literature showing a raise in attentional tunneling. These results were encouraging for trying to automate the diagnosis of attentional tunneling with a formal machine-learning method as ANFIS.

An objective of this study was to test the efficiency of using ANFIS to identify degraded attentional states. The results on the training dataset show that this approach was appropriate to identify the links between attentional tunneling and the psycho-physiological metrics. Indeed, when using the continuous output, the ANFIS matched the checking dataset with an error of 8.9%. When rounded, the ANFIS output matched the checking dataset with 1.1% error. The performance on this task is better than the performance of the SVM, one of the reference techniques in the machine-learning domain.

The classifier also proved to be robust over new data from the four participants kept for testing and could identify the moment when the participants stopped facing attentional tunneling. This promising result shows that it is possible to train the ANFIS on a restricted set of participants who faced attentional tunneling, and then to use this trained ANFIS as an attentional tunneling inference system. Indeed the rules learned from the training were relevant and could be generalized to diagnose attentional tunneling across new participants.

Furthermore, our results suggested that the contribution of the NBAOI metric is most important to the classification, followed by SWR and finally HR. This order is coherent with the observed effect sizes (ETA 2 p) in the inferential analysis (group × segment type interaction). The proportion of the variance attributable to each metrics was ordered identically (NBAOI, ETA 2 p = 0.57; SWR, ETA 2 p = 0.51; HR, ETA 2 p = 0.46). From the human factors point of view, it means that the identification of attentional tunneling could not be optimized without the other two metrics (SWR and HR). In other words, an operator who gazes upon a very limited number of AOIs would not be considered as facing attentional tunneling without demonstrating a low SWR between these AOIs and a high HR.

Another great interest of ANFIS is that it provided objective rules that could be interpreted in terms of natural language. This advantage is particularly relevant for human factor purposes as it allows experts to compare them with known relationships between metrics and cognitive performance. The results of our study revealed that the rules of our classifier were consistent with the literature about attentional tunneling. Indeed, they highlighted that attentional tunneling

was related with a strong reduction of the NBAOI and a decrease in theSWRwhich corresponded respectively to fewer scanned AOI on the user interface and a decreased saccadic activity as proposed by Thomas and Wickens (2004), Cowen et al. (2002). Furthermore, the HR indicator confirmed that this narrowing of the visual field on specific sources was accompanied with psychological stress as previously demonstrated by Bahrick et al. (1952), Easterbrook (1991), Weltman et al. (1966). On top of giving the rules in the natural language, the ANFIS provided numerical thresholds that could be identified while translating the fuzzy domains into the metrics domain.

For these reasons, the ANFIS appeared to be a relevant tool for human factor experts, both as mean to automate the analysis of attentional state metrics in real time and adapt the system to the current attentional impairment ofthe operator, and to better understand which are the key metrics to detect a particular attentional impairment.

6

DEFINITION OF THE EYE STATE IDENTIFICATION ALGORITHM AND VALIDATION

RÉSUMÉ EN FRANCAIS

Le chapitre précédent a montré que l'utilisation des métriques AOI-based couplé à une méthode de classification appropriée permet l'identification en temps-réel de la tunnélisation attentionnelle dans un contexte connu, ici une expérience robotique. Ce chapitre présente une seconde approche au problème de l'estimation des états attentionnels d'un opérateur par l'oculométrie, en s'appuyant sur les métriques directes. L'in-té-rêt des métriques directes réside dans leur indépendance à l'interface utilisée. Dans le cadre d'une utilisation avec un oculomètre mobile, il est alors possible d'utiliser ces métriques dans plusieurs contextes sans calibration. Par ailleurs, ces métriques nécessitent beaucoup moins de traitement que les métriques AOI-based ce qui facilite leur calcul en temps réel.

Comme présenté au troisième chapitre, il existe plusieurs algorithmes d'identification de l'activité oculaires qui permettent le calcul de ces métriques, mais de fortes limitations sur leur performance ont été mise en évidence. Notamment, il est mis en évidence que l'utilisation de plusieurs algorithmes spécifiques pour la détection des fixations, saccades, et clignements peut produire des résultats incohérents: il est en effet possible d'avoir certaines positions de l'oeil qui sont associées à la fois à une fixation et à une saccade en même temps. Ceci est contraire au postulat de Marshall (2000) qui stipule que l'oeil est en permanence dans un seul état à la fois. Le premier chapitre de cette étude consiste donc en la proposition de l'Algorithme d'Identification de Etats de l'Oeil (ESIA) qui répond à ce problème et de sa validation par rapport à d'autres algorithmes connus. Les deux chapitres suivants sont deux cas d'application de cet algorithme pour la détection de la tunnélisation attentionnelle et pour la détection de la désorientation attentionnelle.

L'ESIA est un mix d'un algorithme de détection basé sur la dispersion (I-DT) et sur la vélocité (I-VT). L'analyse en continue de la vitesse oculaire permet notamment la prise en compte des poursuites lisses lorsque que l'oeil sort de la zone de dispersion avec une vitesse faible. L'ESIA est pensé comme une machine à état, dont les transitions sont évaluées à chaque nouvelle mesure de la position de l'oeil. Dès qu'une transition provenant d'un état actif est tirée, alors l'état change. Les conditions de transition sont contrôlées pour assurer la fiabilité de la machine à état et garantir les respect du postulat de

Marshall. L'ESIA produit ainsi une séquence d'état de l'oeil associant à chaque position de l'oeil un état et un seul. En parallèle un fichier consignant tous les changements d'état pendant la mesure est généré. Il contient les caractéristiques principales de chaque événement (date de début, date de fin, amplitude, vitesse moyenne, position initiale, position finale). Il est montré ensuite que le vecteur d'état de l'oeil couplé au fichier des états permettent le calcul de toutes les métriques oculaires, mais aussi de nouvelles métriques. Deux nouvelles définitions du ratio de Goldberg exploitation/exploration sont ainsi proposées. Les performances de l'ESIA face à un I-DT et un I-VT classique sont ensuite ensuite évaluées. Pour disposer d'un jeu de données de référence, le codage manuel de séquence de mesure est effectué. Les 3 algorithmes tournent ensuite sur le même jeu de données et produisent chacun une identification des événement dans l'activité oculaire. Leurs résultats sont comparés. Deux jeux de données artificiels sont égalemement générés en ajoutant du bruit blanc centré gaussien dont le paramètre de dsipersion est fixé à 5° et 10°. Ces jeux permettent de tester la robustesse face bruit de mesure des trois algorithmes. Dans tous les cas de tests, l'ESIA présente des performances au moins aussi bonnes que les deux autres algorithmes. Il résiste notamment bien au bruit de mesure. Par ailleurs, la vocation de l'ESIA est d'être utilisé en contexte écologique avec un oculomètre mobile. Dans ces conditions, de nombreuses poursuites lisses sont présentes dans le signal est doivent être correctement identifiées en tant que fixation. Deux jeux de données sont utilisées pour cette évaluation, présentant des différences de nature de poursuites lisses: le premier jeu est enregistré face à un écran d'ordinateur, alors que le second est enregistré dans un cockpit d'Airbus A330. On attend donc que les mouvements de têtes soient plus important dans le second jeu de données que dans le premier. Dans les deux conditions, l'analyse est faite sur la position de l'oeil avec l'environnement comme référentiel de référence d'une part et avec le visage du participant d'autre part. Une étude statistique permet de vérifier que l'ESIA produit des analyses équivalentes que le mouvement de la tête soit pris en compte ou pas ce qui valide son utilisation en contexte écologique avec un oculomètre mobile. Suite à cette validation de l'ESIA en tant qu'algorithme d'identification des états oculaires en condition opérationnelle, deux cas d'étude sont présentés dans les deux chapitres suivants.

ENGLISH VERSION

As presented in the previous chapter, attention is tightly linked with vision, and especially attentional tunnelling which can be detected using ocular metrics as NBAOI and SWR. These attentional metrics that are *interface-dependent* because they are based on AOIs. But using AOIs is incompatible with our goal of providing metrics that do not depend on the interface. Our objectives in this chapter are to propose a mean to compute attentional metrics, in real-time, and which are context independant. In order to achieve such a goal, several limitations had to be overcome.

Avoiding AOIs requires to rely on classical event identification algorithm to analyse the ocular activity. The current event detection algorithms have are not reliable as is (see Chapter 4). The *lack of control* of the threshold was reported as having a dramatic impact on subsequent analysis (Karsh and Breitenbach, 1983, Duchowski, 2007, Shic et al., 2008, Nyström and Holmqvist, 2010, Holmqvist et al., 2011). Furthermore Saccades, fixations and blinks are generally identified independently by different algorithms. Each identification algorithm uses its own criterion for detecting events. As they run independently, there is no cross-check between them. In that configuration, the eye could be identified as being in the state of fixation by the fixation identification algorithm, while also identified as being in the state of saccade by the saccade identification algorithm. This phenomenon is called overlapping, and violates the simple principle of the eye being in one state at a time (Marshall, 2000). It is illustrated on Figure 35.

Our constraints are therefore the following in order to meet our objectives while overcoming the limitations of the current algorithms:

- Propose a new algorithm on which we would have the total control on both the identification strategies and the choice of the detection thresholds.
- Make this algorithm interface independent. In other word, it must handle eye-in-head eye position sequence for performing the event detection.
- Make sure that the algorithm respects the non-overlapping principle.

The Eye State Identification Algorithm (ESIA) is our proposal for achieving such objectives. It associates one state with each ocular measurement from the eye-in-head position. This "eye state" approach was suggested by Marshall (2000) and offers new opportunities of metric computation, on top of the classical metrics that the ESIA also produces. We will first explain the states that we would like to detect with the ESIA before presenting the algorithm itself and the derived metrics in details. Its performance is then compared to two reference

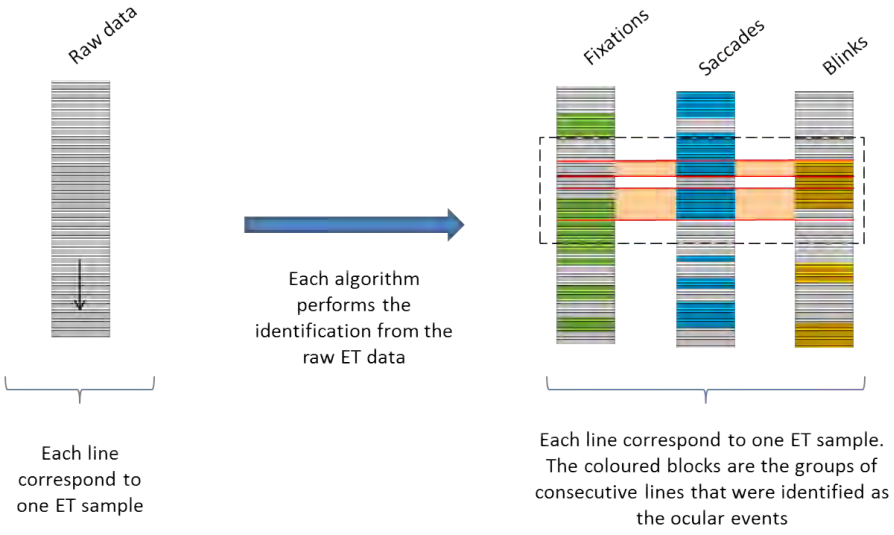


Figure 35: Examples of the event identification limitations. In the black dotted rectangle on the right, the red zones identify samples that are identified as being in fixation and blinks at the same time, thus violating the non-overlapping principle.

identification algorithms. Then we will demonstrate that using the eye-in-head and eye-in-the environment is equivalent for the ESIA. Finally, we will show that the ESIA can be used to estimate interface-dependent metrics such as the SWR without any knowledge of the interface.

In this chapter, we present an alternative approach for computing interface-independent attentional metrics on line.

6.1 THE EYE STATES

This subsection was presented in an precedent version in (Regis et al., 2012) and reuses results from Chapter 4. The states that are identified by the ESIA are of 3 types: *blinks* that occur when the eyelids cover the eye globes, *saccades* that are the rapid ballistic movements of the eye between two fixations with high velocity, and *fixations* that are the relatively stable positions of the eye allowing for information encoding. The ESIA is an aggregation of both I-DT and I-VT detection algorithms for including smooth pursuit (see Chapter 4) and finally detecting blinks, saccades and fixations robustly. It also includes two additional modules that classify fixations into 4 subtypes (express, short, medium and long), and saccades into 2 subtypes (small and large).

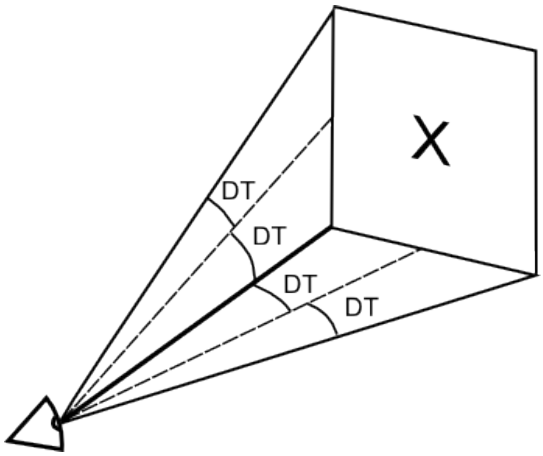


Figure 36: Representation of the area of fixation depending on DT

6.1.1 *Blinks*

Blinks are considered as interrupting any ocular events and are therefore always considered before searching for saccades or fixations. Only if no blinks are detected is the eye position sequence analysed for extracting fixations and saccades. Most of the time, blink detection is performed by the eye-tracker itself via a dedicated algorithm, or as a specific signature in the position signal. No norm can be found on blink detection among eye-tracking manufacturers, but most of the time, a blink binary sequence is provided. If not, an additional blink identification algorithm could provide such data. For that purpose, we have integrated this binary sequence as an input in the ESIA.

6.1.2 *Fixations and saccades detection*

Fixations are defined as relatively stable positions of the eye, with a reference dispersion value of 1°. The ESIA uses a dispersion-based approach to detect fixations, similar to an I-DT. The dispersion threshold (DT) is the same for both the vertical and horizontal dimensions, which means that the fixation area is a square whose sides length is 2 DT degrees (see Figure 36). .

As explained in Chapter 2 on the Human Visual System, smooth pursuits and fixations involve the same network that is responsible for maintaining the eye on the target of interest. The ESIA therefore identifies smooth pursuits as fixations. If the eye goes out of the fixation area but with a velocity that is under the velocity threshold VT, then the eye is considered as being in the fixation state. The reference VT parameter in the literature is $30° s^{-1}$.

The VT and DT thresholds of the ESIA can be chosen at will depending on the specific requirements of the experiment. They can be

set in pixels or in degrees. The parameters proposed here have been inspired from previous studies as presented in Chapter 2.

6.1.3 *Additional classification for saccade subtypes*

Visual psychophysics suggest a segregation of the visual field into the foveal/parafoveal visual field (less than 4-5° of eccentricity) and the peripheral visual field (more than 5° of eccentricity). At the 5° boundary, the visual acuity is already divised by 2 (Irwin, 1992) Furthermore, as a fixation could be classified as focal or ambient, Follet et al. (2011) suggested that the classification could be also applied to saccades. In the context of free viewing, they segregated two clusters of saccades, one of small extent saccades categorized as focal (cluster centred at 3°), and one of large extent saccades categorized as ambient (cluster centred at 11°). Saccades are stereotyped movements of the eye, with an almost bijective link between the saccade amplitude and the peak velocity (Bahill et al., 1975). Saccades of 4-5° have a peak velocity of 200-300 $^\circ\,s^{-1}$. Inspired by these authors, we propose the following saccade classification:

1. Saccades with a peak velocity under 250$^\circ\,s^{-1}$ are classified as *small, focal* saccades.
2. Saccades with a peak velocity above 250$^\circ\,s^{-1}$ are classified as *large, ambient* saccades.

Though the threshold values have been recommended here, the threshold have been implemented as parameters of the classification and can be set at will. This allows for fine tuning the classification to a specific participant or a task for instance.

6.1.4 *Additional classification for fixation subtypes*

Depending on the fixation length, we proposed a fixation classification in four types. The literature indeed associates different roles depending on the fixation duration. Though authors tend to discard fixations that are shorter than a certain period (80 ms - 100 ms), Velichkovsky et al. (2000) proposed to label them as "express" fixations. Graf and Krueger (1989) suggested that fixations shorter than 240 ms are involuntary, whereas fixations longer than 320 ms are voluntary. Also Velichkovsky et al. (2005) introduced focal versus ambient fixations that would relate the visual pathways activity. Focal fixations would be linked to the temporal pathway for encoding and making sense of information at the centre of the visual field, whereas ambient fixation would be linked to peripheral processing via the ventral pathway. These authors classified ambient fixations in the 90 - 260 ms domain, and focal fixations in the 280 ms and above domain. We have used these models for proposing our own temporal classification:

1. Fixations shorter than 80 ms are classified as *express* fixations (consistently with Velichkovsky et al. (2000)).
2. Fixations between 80 ms and 240 ms are classified as *short, involuntary or ambient* fixations (consistently with Graf and Krueger (1989), Velichkovsky et al. (2005)).
3. Fixations between 240 and 320 ms are classified as: *medium, indetermined* fixations (consistently with Velichkovsky et al. (2005) who used a 280 ms threshold, and Graf and Krueger (1989))
4. Fixations longer than 320 ms are classified as *long, voluntary or focal* fixations (consistently with Graf and Krueger (1989), Velichkovsky et al. (2005)).

Though thresholds values have been recommended here, the thresholds have been implemented as parameters of the classification and can be set at will. This allows for fine tuning the classification to a specific participant or a task for instance.

6.1.5 *Computed eye states*

Though interesting interpretations of saccades and fixations being focal or ambient, either voluntary or involuntary have been proposed, we have decided to keep neutral names as this topic is still debated in the community. Finally, the ESIA can identify 3 eye states (blink, saccade, fixation), with 2 saccades subtypes and 4 fixations subtypes :

- Blink (coded 0)
- Saccade :
  - Small (coded 1)
  - Large (coded 2)
- Fixation :
  - Express fixation (coded 3)
  - Short fixation (coded 4)
  - Medium fixation (coded 5)
  - Long fixation (coded 6)

6.2 THE EYE STATE IDENTIFICATION ALGORITHM

The first version of the ESIA was published in the proceedings of the Human Factors and Ergonomics Society Chapter Europe conference (Regis et al., 2012). Its latest version is detailed in this section.

6.2.1 *Inputs*

As explained in the introduction section, our objective is to consider the the eye-in-head orientation of the eye. The most appropriate sensors for providing such information are the head-mouted eye-trackers (see Chapter 4). The ESIA inputs are nevertheless generic for all eye-trackers: Timestamp, Eye X, Eye Y, and Status. They correspond respectively to the sample recording time, the horizontal and vertical eye-in-head orientation (relative to the head), and the pupil detection status (that takes the value 0 if the pupil is not detected, and 1 otherwise). In the following algorithm description, the Status null-value is considered as being due to a blink for simplification. The data source used as input (the eye-in-head position) does not take the context into account in the computation process. All derived metrics will therefore be *interface-independent*.

6.2.2 *Algorithm for spatial classification*

The ESIA is an implementation of the state machine that is represented in Figure 38. In this representation, a token can transit from place to place (the squares and circles) through the transitions (arrows). A place is marked, or active, when the mark is on that place. A transition is called active when the conditions are met for allowing the transition between two places.

The state is updated following each new eye-position sample provided by the eye tracker. It is synchronous with the eye-tracker's clock, so state analysis is performed *on line*. Each new sample allows for the conditions of all transitions going out of the current active place to be evaluated. The conditions have been chosen for allowing only one transition to be active at a time, therefore resulting in only one place marked at a time. This guarantees to tag one and *one eye state only per sample*. The state machine approach can be compared to the two-states I-HMM detection algorithm as presented in (Salvucci and Goldberg, 2000). First, I-HMM involves only two states (saccade and fixation), against 3 here (saccade, fixation, blink). Second, transitions are performed on probabilistic basis with I-HMM whereas with the ESIA, transitions are based on deterministic conditions based on DT and VT. The next subsection simulates the state machine functioning, helping the reader to understand the state machine representation of the ESIA (see Figure 38).

6.2.3 *A simulation of the state machine*

1. At the beginning, the INIT place is marked.

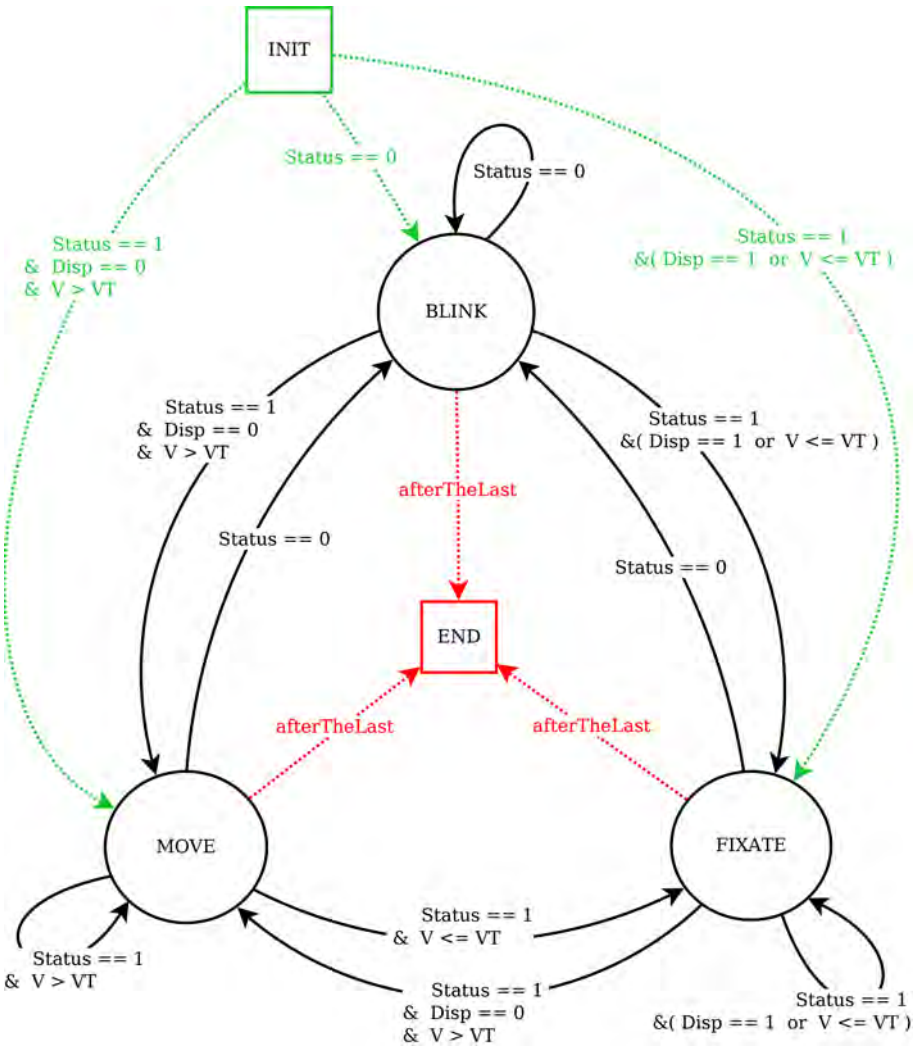


Figure 37: State machine representation of the ESIA. The "afterTheLast" transition is given priority on any other transition.

2. When the first sample arrives, Status value is tested. If it is 0 ($Status == 0$). The transition to BLINK becomes active. The token goes from INIT to BLINK. It is consistent with the hypotheses that when the pupil is not detected it is due to a blink.

3. A new sample arrives. The active state is BLINK, therefore the conditions of all the output transitions are evaluated to figure out which is the active transition. Status is first tested. If it is 0 again, then the machine remains in the BLINK state. Otherwise, Disp is tested. Disp is 1 if both consecutive positions are inside the dispersion zone fixed by DT. In that case, the transition to FIXATE place becomes active, and the FIXATE place is marked. The position of the first sample is stored in a buffer and *considered as the center of the dispersion zone*.

4. Let us consider that among the following samples, the first ten have Status value at 1, and all fall in the dispersion zone. The state machine will remain in the FIXATE state for 10 clocksteps. Now, the next sample falls out of the dispersion zone, but Status is still at 1: the velocity condition is therefore evaluated. If velocity is under VT, the FIXATE to FIXATE transition is active again (according to the conditions), and the FIXATE place remains active. This procedure allows for smooth pursuit to be integrated into the FIXATE state. The *centre of the dispersion zone is updated* to this new eye position.

5. The dispersion zone has been updated. Let us consider that the next few samples fall in the dispersion zone, or, if they fall out the dispersion zone, they have a velocity that is under VT. Then the FIXATE place will remain active, and the dispersion zone centre will be updated as many times as required.

6. One new sample arrives. If its Status is 0, then the mark goes to BLINK. FIXATE block is over. The duration of the fixation has been incrementally calculated. If Status is 1, but the sample falls out of the dispersion zone *and* the velocity is higher than VT, then the mark goes to MOVE. It will stay in MOVE as long as the velocity is above VT and Status is 1.

7. The process goes on, with mark evolving from place to place, until the end of the sequence is reached. The end of the sequence is detected via a sample that has -1 values on each field. This activates the "afterTheLast" condition. Note that this transition is in fact evaluated with each new sample, and has priority on any other conditions for activating the transition to the END place.

The algorithm therefore starts in the INIT state, marks all new sample with BLINK, FIXATE or MOVE, and ends in the END state when

no more samples are provided. While performing the analysis, the fixation duration is analysed in order to determine the fixation type to be tagged.

6.2.4 *Outputs*

The ESIA outputs a tag for each eye position sample. This series of states is called the *Eye State Vector (ESV)*. Please see Figure 38 for a representation of the ESV on eye-tracking data.

The ESIA also computes a *log* of the detected events that are stored in a text file in which each tab-separated line corresponds to an event. Each log line contains the following fields: event type (integer), start time (double), duration (double), coordinates of the first and last positions of the event (4 doubles), extent of the event (distance between the first and last position of the event) (double), and the maximum velocity of the event (double). These fields are calculated whatever the event though the extent and maximum velocity are more suited for saccades, and duration is more suited to fixations and blinks.

6.3 BASIC METRICS DERIVED FROM THE ESIA

The ESIA allows for basic metrics to be computed. The log is of special interest for that purpose. From this log several metrics can be computed as presented in this section.

6.3.1 *Example of basic metrics calculation on a period of interest*

In the course of an experiment, some periods are generally identified as the period of interest (POI). These periods are the rest period for instance, or few seconds before and after an experimental event such as a stimulus onset. POI can indeed be linked to experimental event markers (triggers). For instance, an alarm generates a trigger that identifies the exact time of the alarm onset. An example of a classical POI would be the 30 seconds following the alarm onset.

Isolating the events belonging to any period of interest (POI) using the start time is straightforward. The log of this subset of events still contains most of the data that are used for basic metrics computation such as:

NUMBER OF FIXATIONS: counts the number of events in the subset whose event types are fixations of all types ($eventtype == 3, 4, 5 or 6$). This works for saccades, and blinks.

AVERAGE FREQUENCE OF FIXATION DURING THE POI: divide the number of fixations by the length of the POI. This works for saccades, and blinks.

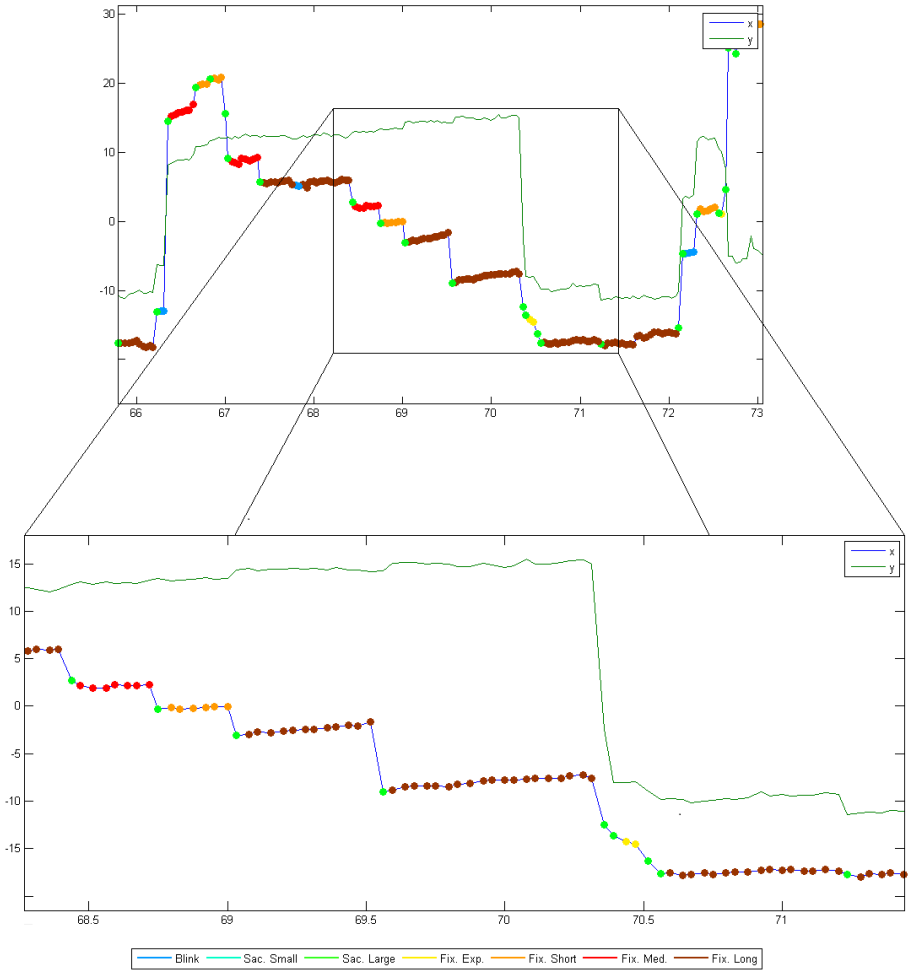


Figure 38: Graphical representation of the ESV output. X-axis is in seconds, whereas Y-axis is in degrees. The blue and green solid lines represent respectively horizontal, and vertical eye-in-head position series. ESV tags are superimposed on the horizontal eye position series. Green tags are saccades, yellow-orange-red and maroon are express, short, medium, and long fixations, blue tags are blinks.

MEAN FIXATION TIME: isolates all fixations and extract their duration for computing the mean. This works for saccades, and blinks.

TOTAL DISTANCE DURING THE POI: sums the extent value of all events of the subset.

MEAN PEAK VELOCITY: isolates all saccades and extracts their maximum velocity for computing the mean.

These are just a few examples showing that classical metrics presented in Chapter 4 are accessible with the ESIA thanks to the log. But state tagging allows for other metrics to be computed, which that are of special interest.

6.3.2 *Heat maps and Scanpaths*

Heat maps and scanpaths can also be computed easily from the log. The centroids of fixations are provided together with their duration. The formulas for computing heat maps from fixations centroids has been given in Chapter 4. The coordinates of the first and last positions of each event is stored, which facilitates the computation of scanpaths (they are the sequence of the event vectors defined by the end coordinates minus the start coordinates).

6.3.3 *Focal versus Ambient*

The key in studying focal versus ambient fixation is to extract all fixation-saccade couples (Unema et al., 2005, Velichkovsky et al., 2000, 2002, 2005). These authors advocated for a couple to be formed with a fixation and the subsequent saccade. But recently, Follet et al. (2011) advocated for a classification using the previous saccade instead of the subsequent. In fact they did not construct couples but trios of saccade-fixation-saccade (see Figure 39).

The ESIA allows for computing such couples or trios from the log. Most of the log lines involve a succession of saccades and fixations, with only rarely a blink. Building fixation-saccade couple with the log is easy: each time the event type is $3, 4, 5$ or 6 (meaning that it is a fixation), then a simple test on the next event checks that the next event is a saccade or not with the event type. If it is, then a couple is formed, via the concatenation of the two lines of the log for instance. Any combination of saccade-fixation can be built this way. It allows for further analysis as it will be demonstrated in Chapter 8.

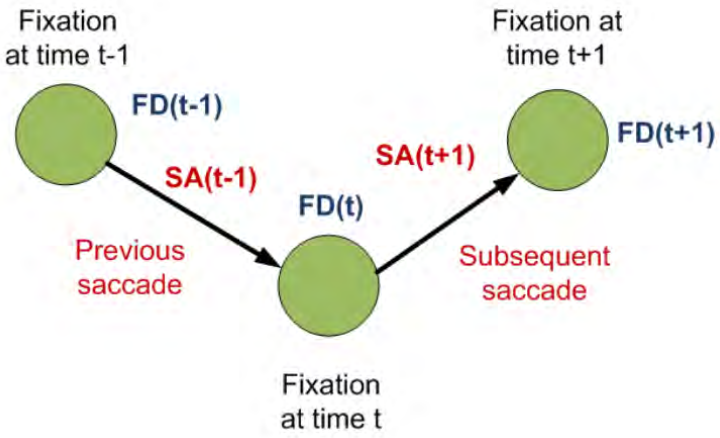


Figure 39: Saccade-Fixation-Saccade Trio for the study of focal versus ambient fixations. From (Follet et al., 2011)

6.3.4 *State distribution*

The time spent in one particular state during any window is computed instantaneously with the ESV. This measurement was proposed in (Regis et al., 2012). The sum of the time spent in all the possible states is equal to the exact duration of the window. This allows for a relevant metric to be computed, i. e. the state distribution.

The state distribution is the relative contribution of all states in the window. For each state, it is the number of samples in that particular state divided by the total number of samples in the window. It results in a value that is between 0 and 1. The relative contribution of all states in the window sum to 1, whatever the window.

The state distribution was computed using the last 10 seconds of records for all participants of the "Computer" experiment. As a consequence, each participants was assigned 7 series, consisting of the share of the ocular resource occupied by each state at any time in the experiment. An example of a state distribution is presented on Figure 40.

6.3.5 *Correlation between states*

As presented in the previous section, the state distribution sums to one and changes in the relative shapes of the distribution should reflect the current ocular optimization. In order to better understand the relative evolution of such series, a correlation analysis was carried out. The 7 series from all participants were concatenated to form 7 bigger series covering the whole experiment and finally consisting of 101533 samples each. Two additional series were formed: one as the sum of the two saccades subtypes series, and one as the sum of the four fixation subtypes series. We evaluated the correlation of

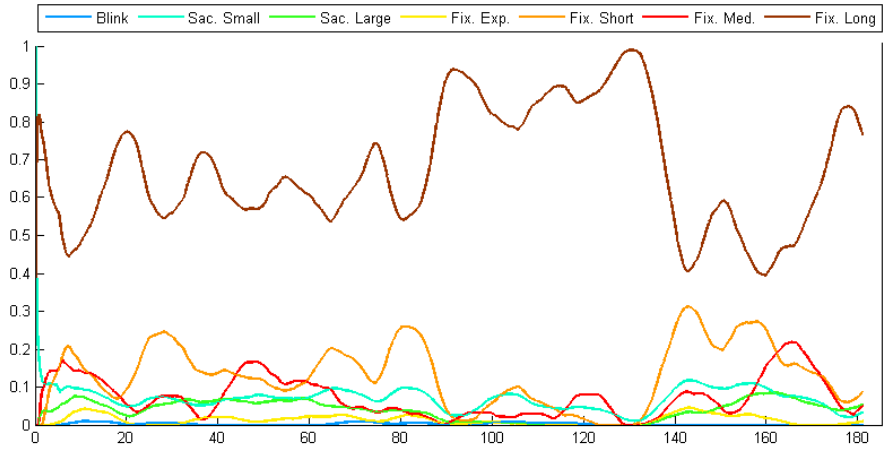


Figure 40: State distribution computed from the ESIA. For each second of the experiment, the state distribution was calculated on a moving POI consisting in the last 10 seconds. Filtered with a 5 sec. moving average. Express fixations were counted as saccades. Each curve representes the relative contribution of one state, with a different color for each as stated in the legend. Curves sum to one at any instant of the record.

these 9 series over the 101533 samples with a multi-linear regression analysis. All correlation were significant (maximum p-value of 10^{-14}, which is consistent with the number of samples). Results of the correlation analysis are given on Figure 41:

- Blinks (0) are not correlated with saccades $(1+2)$, but negatively correlated with fixations $(3+4+5+6)$. The longer the fixations the more negatively correlated, though the correlation coefficients remain close to 0.

- Saccades $(1+2)$ are positively correlated with fixation $(3,4,5,6)$. The longer the fixation the less correlated. Saccades $(1+2)$ are negatively correlated with long fixations (6) with which they show strong negative correlation.

- In detail, small saccades 1 show a strong positive correlation with express fixations 3 decreasing with short (4) and medium (5) fixations, and negatively correlated with long fixations (6).

- Large saccades 2, is positively correlated with express (3), small (4), and medium fixations (5).

- Fixations $(3+4+5+6)$ are negatively correlated with blinks (0), and saccades $(1+2)$.

- In detail, Express (3) and short fixations (4) show strongest correlation with saccades $(1, 2, 1+2)$.

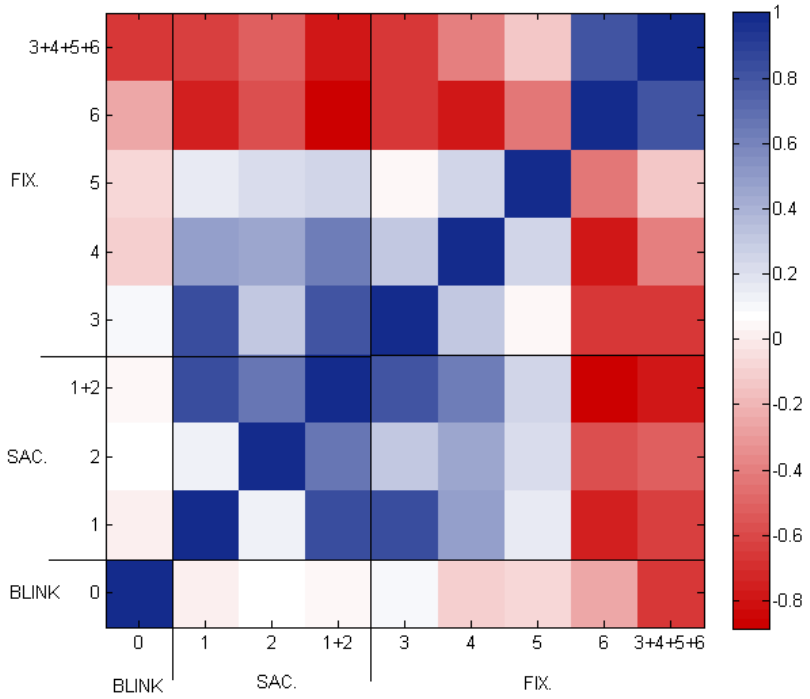


Figure 41: Correlation maps between ocular metrics from the state distribution. Each coloured square represents the correlation coefficients as presented on the colour map on the right of the figure. For instance Medium fixations (number 5 on the X axis) has a correlation close to 0 (white square) with express fixations number (number 3 on the Y axis).

- Medium fixations (5) are not clearly correlated with any state.
- Long fixations (6) are negatively with all other events (0, 1, 2, 3, 4, 5).

These results are consistent with what is known on the ocular activity: as blinks interrupt the visual intake, whereas long fixations are intense encoding, it seems natural that blinks tend to disappear with longer fixations. Saccades are correlated with all types of fixation, especially short and express ones and negatively correlated with long fixations, which is consistent with the ambient versus focal classification of fixations: indeed short fixations would be associated with a higher saccadic activity for information search. Finally, long fixations are negatively correlated with all other states. This is consistent with the information encoding strategy which requires long fixations that leave little or no room to the other states.

As a summary, two groups can be identified from this correlation analysis: 1) saccades and all but long fixations for visual search 2) long fixations for in-depth information encoding on the other hand. This grouping is slightly different from what our temporal classification would have suggested that is: 1) large saccades, express and short fixations for visual search 2) small saccades and long fixations for information encoding.

6.3.6 *Exploit versus Explore Ratio*

The state distribution correlation analysis raises valuable insights on the ocular resource optimization strategy. Two modes are indeed elicited: information encoding with long fixations, and visual search with saccades and short fixations.

To some extent, these two modes correspond to solutions of the optimization problem of maximising the outcome of multiple sources generating unknown gains (the environment stimuli), under the constraint of being limited to exploit only one source at a time (like the eye), with costs associated with shifting from a source to another (shifting costs as presented in Section 3.3.3). This optimization problem results in a trade-off between *exploiting* continuously the same source (at the cost of possibly neglecting some others with greater outcomes), or *explore* for other sources (at the costs of shifting, and possibly find only less productive sources). In the context of task supervision, neural correlates of this type of optimization in the brain has been presented by Aston-Jones and Cohen (2005). He studied the role of the Locus Coeruleus-Norepinephrine (LC-NE) in optimizing overall task performance and suggested two distinct modes of functioning:

- "The LC tonic mode produces adaptive adjustments of gain that optimize performance *across tasks* (exploration versus exploitation)".

- "The LC phasic mode produces adaptive adjustments of gain that optimize performance *within a task* (temporal filtering)"

The purpose of introducing this analogy is not to demonstrate a direct link with the ocular activity, which would require much more arguments. Nevertheless, Goldberg and Kotval (1999) proposed an ocular counterpart of such an analogy and defined the "exploit versus explore ratio", an ocular metric that was presented in Section 4.4.1 and whose calculation from the ESIA can be found in Section 6.3.4.

Goldberg suggested one definition of the exploit/explore ratio. The correlation analysis and the temporal classification suggested two other definitions that are presented below. All ratios used the previously defined state distribution, calculated on a 10-second window. For all ratios R_{xxx}, a z-score normalisation was applied for inter individual comparison named R_{N-xxx}. For all definitions FIX(subtypes) denote the subtypes of fixations in the parentheses. Idem for SAC(subtypes) for saccades. The first ratio definition R_{gold} respects the definition found in Goldberg and Kotval (1999), and is here adapted to the ESIA outputs:

$$R_{gold} = \frac{Exploit}{Explore}$$

$$R_{gold} = \frac{FIX\,(Exp + Short + Med + Long)}{SAC\,(Small + Large)} \tag{5a}$$

$$R_{N-gold} = zscore(R_{gold}) \tag{5b}$$

The second ratio definition R_{correl} is compliant with the state distribution correlation analysis conclusion in Section 6.3.5 in this chapter:

$$R_{correl} = \frac{Exploit}{Explore}$$

$$R_{correl} = \frac{FIX\,(Long)}{SAC\,(Small + Large) + FIX\,(Exp + Short + Med)} \tag{6a}$$

$$R_{N-correl} = zscore(R_{correl}) \tag{6b}$$

The third ratio definition R_{ESIA} is compliant with the additional temporal classification proposed with the ESIA (see Section 6.1.4):

$$R_{ESIA} = \frac{Exploit}{Explore}$$

$$R_{ESIA} = \frac{FIX\,(Long) + SAC\,(Small)}{SAC\,(Large) + FIX\,(Exp + Short)} \tag{7a}$$

$$R_{N-ESIA} = zscore(R_{ESIA}) \tag{7b}$$

When not normalized, the ratios can be computed on-line. The mean and the standard deviation are calculated on a period of the scenario. If this period is the whole scenario, then the calculation is

not possible on-line anymore because samples at the end of the experiment are required. These ratios were tested on the three use cases that are presented afterwards.

6.4 BENCHMARK

Surprisingly, benchmarking is not a common practice in eye-tracking. As presented in Chapter 4, the community is faces a lack of standardization of the event identification algorithms. This comes from the eye-tracking industry that provides eye-trackers with their own identification algorithms and filtering policy and give little information about them, but also from the various definitions proposed by the researchers themselves and the strong impact of the differences between algorithms (Karsh and Breitenbach, 1983). Furthermore, there is no eye-tracking dataset available with saccades and fixations events certified by experts. Evaluating the performance of a new identification algorithm is therefore an issue by itself.

Nyström and Holmqvist (2010) proposed an adaptive algorithm for saccades, fixations and glissades detection including a benchmark of their algorithm. But they still faced the limitations that were just presented: identification algorithms were compared against each other without a reference, and their adaptive algorithm based on Savitsky-Goulay filtering was compared against I-DT and I-VT without any filtering, with no rejection of non-physical measurement. For the benchmarking of the ESIA, we have coded manually a dataset that has been used as a reference and we have compared performance with I-DT and I-VT. We have also artificially generated two datasets from the original dataset with added noise on the eye position sequence.

6.4.1 *I-DT and I-VT definition*

We have implemented I-DT and I-VT according to their most common definition:

I-DT considers each sample as a potential fixation. The dispersion condition is then evaluated between the first sample position and the succeeding ones. As soon as the sample position is out of the dispersion area (the distance between the first position and the current on both axis exceed DT), the fixation block is stopped (it can be one sample long only).

I-VT considers each sample as a potential saccade. The velocity condition is evaluated at the first position, and if the velocity is under the velocity threshold VT, then the sample is tagged as being a fixation and the next position is considered. Otherwise, the first position is considered as a saccade, and the next sam-

ples are associated with the saccade as long as their velocity is above the threshold.

6.4.2 *Source data and manual coding*

We coded manually $10 * 10$ seconds of eye-tracking data at 100Hz (10000 samples) in order to build the reference dataset. These data were extracted randomly from recordings made with the fixed and constrained EyeLink1000+, sub-sampled at 100Hz, on 10 trainee pilots performing a psychomotor test on a computer screen for evaluating their skills in multitasking management.

This dataset was selected because of the high quality of the data, therefore the manual identification was straightforward. We determined expertly which periods of the recordings were saccades (manual coding) and tagged every consecutive samples of these periods with a "1". The remaining blocks of "0" were considered as fixations. We categorized the fixations (blocks of consecutive zeros) as express, short, medium and long according to the temporal parameters proposed in the previous section. We also generated two datasets from this reference with gaussian noise (with σ of 5 and 10 pixels) that was added to the reference signal. The goal was to test the robustness of the algorithms on noisy data, which are common in eye-tracking (see Figure 42).

6.4.3 *Event identification and comparison*

The ESIA, I-DT and I-VT were run according to their respective definition in the Matlab2013a environment. We used the same spatial parameters for all algorithms in order to compare their performance: DT=30 pix and VT = 2000 pix s^{-1}(equivalent to a dispersion area of 2°and a velocity of 67° s^{-1}). In order to compare the output of the three algorithms, we had to recode the sequence produced by the I-DT and the I-VT and the manually coded sequences. In other words, we categorized the fixation subtypes according to the temporal parameters, and then re-tagged all the fixations subtypes of both sequence. We did not differentiate the saccades at that stage. The equality between the reference tags and the tags identified by the 3 algorithms could therefore be evaluated. Performance was calculated in terms of percentage of identical tags on the whole dataset (see Figure 43a), and in terms of number of events found in the log (see Figure 43b).

All algorithms performed well on the original data with no noise. Especially, the I-VT, despite its simplicity, has performance that compares well with the ESIA, with the advantage of being simpler. Nevertheless, adding noise on the measurements ($\sigma = 5$pix) also adds velocity noise which the I-VT is very sensitive to. It therefore detects

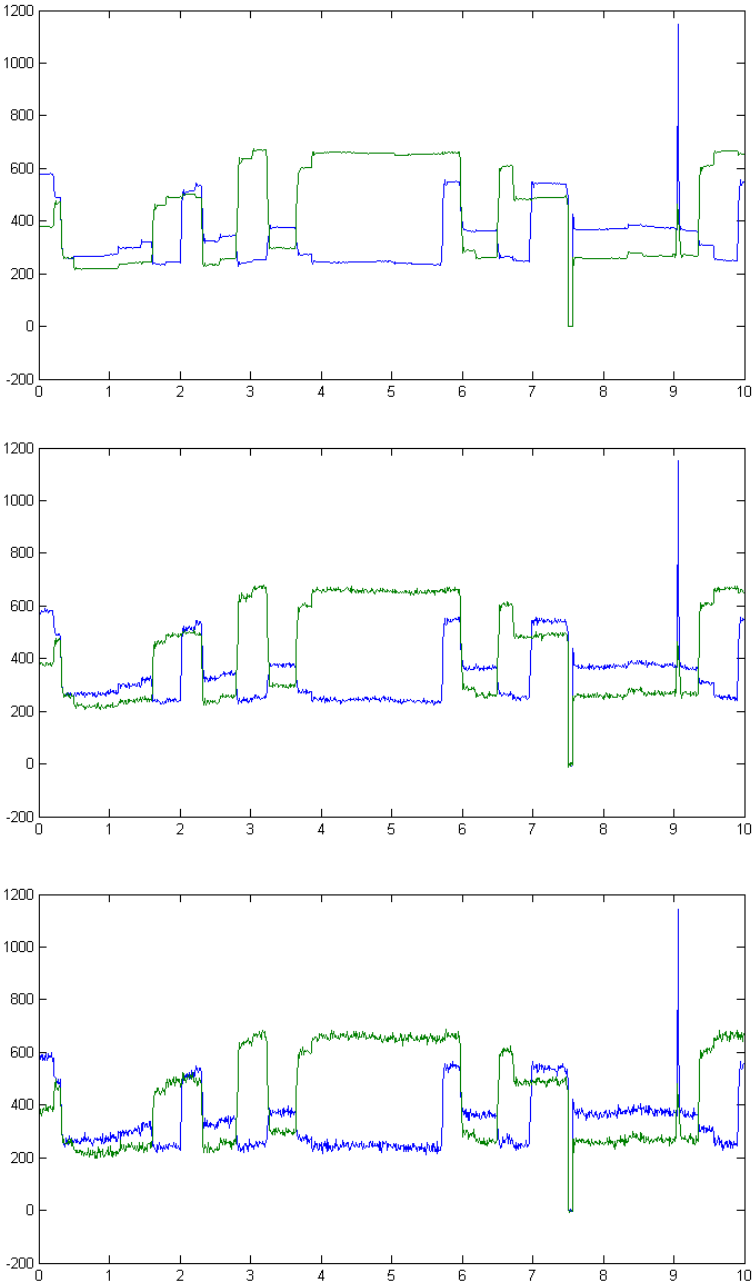
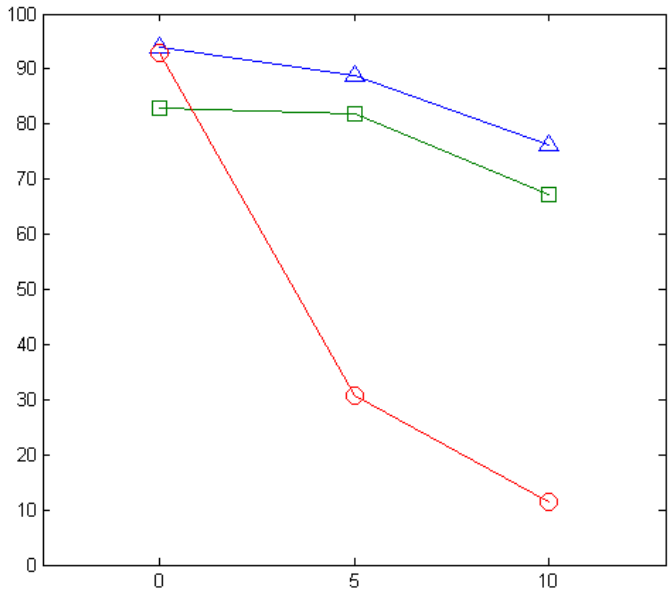
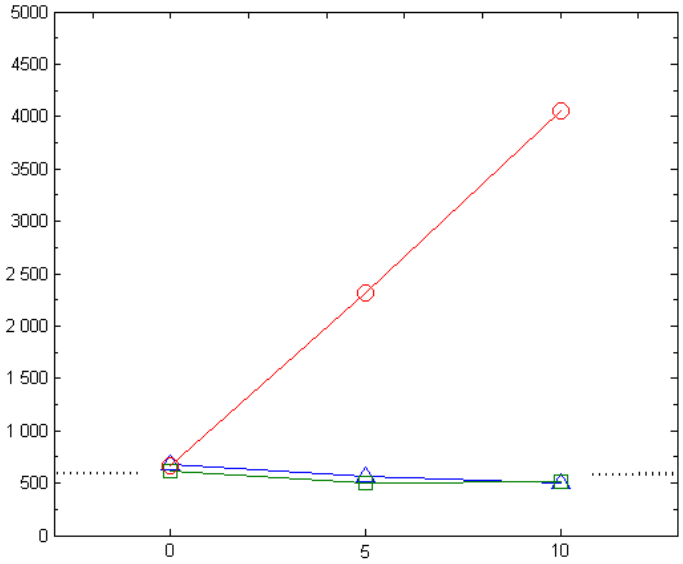


Figure 42: Example of source data used for benchmarking. This corresponds to 10 seconds of records. From the top to bottom: no noise added, white noise with $\sigma = 5$pix , white noise with $\sigma = 10$pix. Horizontal position is plotted in blue, whereas vertical position is in green (in pixels relative to the computer screen).



(a) Percentage of samples tagged correctly.



(b) Number of events detected. Dotted lines show the right value from the manual coding

Figure 43: Comparison of performance between ESIA, I-DT, I-VT. Blue triangles for the ESIA, green squares for the I-DT, red circles for the I-VT.

false saccades because of the high speed between consecutive samples whereas the I-DT still detects these samples as fixations because they all fall in the same fixation area. For $\sigma = 10$pix, the noise distribution is wide enough to impact even the I-DT algorithm. Both ESIA and I-DT show a drop in performance: the measured signal is too noisy for the chosen thresholds. The number of events detected by the I-DT and the ESIA diminishes, which is contrary to our expectations. But the number of events detected by the I-VT rises as the velocity noise triggers more and more saccades. On average ESIA performs better than the other two algorithms on this dataset.

6.5 INFLUENCE OF THE HEAD MOTION IN THE ESIA IDENTIFICATION

One major concern when using eye-tracking in realistic situations such as flying an aircraft or driving a car, is the presence of head movements that accompany the eye for reaching the furthest target in the peripheral field of view. As presented by Wickens et al. (2012), eyes generally move alone for targets that present less than 20° of eccentricity, head is mobilized when the eccentricity exceeds 20°, and above 90°, the whole body rotates. Head-motion will therefore impact the eye-tracking measurement.

We wanted to focus on the eye-in-head ocular activity. But fixating an object while the head is moving induces eye rotations that are exactly the opposite of the head movements. These eye movements are measured in the eye-in-head position sequence. This problem is equivalent to fixations on moving targets while the head is fixed (because of the velocity composition law), which typically happens on dynamical stimuli like a video on a computer screen.

One constraint of the ESIA is to handle to smooth pursuits in order to be suitable for realistic measurement with head motions. Our hypothesis was the following : "The ESIA should result in similar identification, whether the head motion is taken into account or not".

In order to account for the impact of head movements in identifying the ocular states with the ESIA, we tested two datasets in which both eye-in-head positions, and eye-in-environment positions were computed in parallel. We ran the ESIA using these two series and compared them for similarity.

6.5.1 *Datasets*

No specific experiment was designed for testing our hypothesis. We used 2 datasets recorded in two experiments with different experimental conditions. We selected the datasets with one involving almost no head movement, contrary to the other. The head motion was recorded parallel to the eye motion in both datasets.

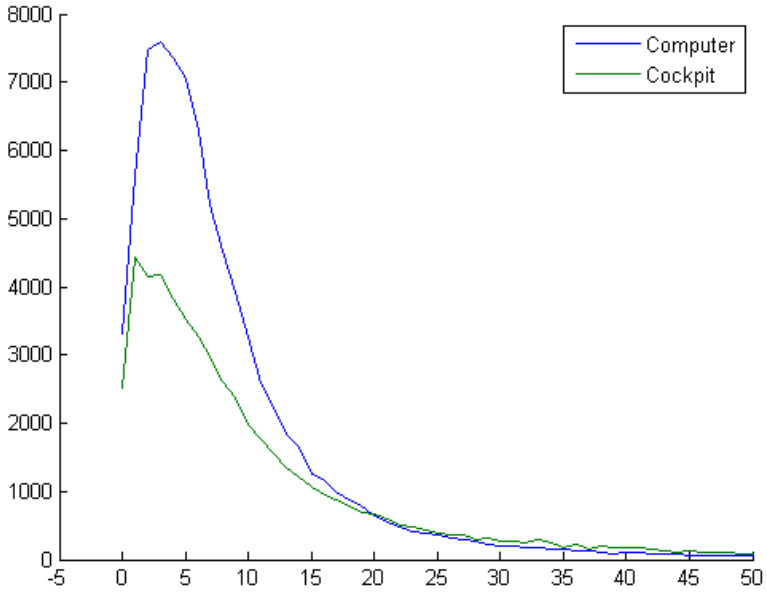


Figure 44: Speed histogram of the "low head motion" (top) and "high head motion" (bottom) datasets, expressed in $°s^{-1}$. Right extremity corresponds to the 99% quantile.

The first dataset comes from the experiment presented in the first contribution in Chapter 5. Participants were supposed to supervise a semi-autonomous robot whose interface was displayed on a computer screen. According to Wickens et al. (2012), we expected the head to remain fixed most of the time because the whole screen is included in a cone of less than 20°of eccentricity. This dataset was our low head motion dataset, called "Computer" dataset in the subsequent parts. It is composed of 82195 samples, recorded at 25Hz with the head-mounted EyeTechSensor from Pertech.

The second dataset comes from an experiment carried out in a full flight simulator of a commercial airliner (A330). Because displays and buttons are much more spread out than in the other experimental conditions, we expected the participants to have much more head motions. This dataset was our high head motion dataset, called "Cockpit" dataset. It is composed of 56366 samples., recorded at 50Hz with the head-mounted EyeTechSensor from Pertech, and sub-sampled at 25Hz.

6.5.2 *Head velocity distribution*

We evaluated the speed distribution over all participants in both experiments. The first step was to verify that the experimental conditions indeed involved different types of head motions so that we

could compare them. We therefore studied the head velocity distribution for these two datasets.

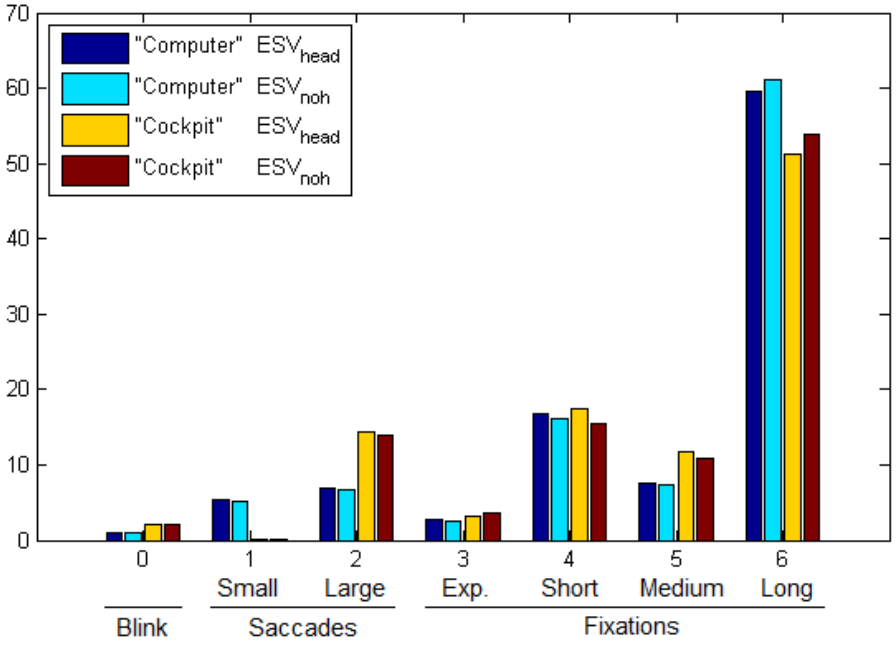
The histograms for both distributions Figure 44 present similar profiles. A t-test was run and confirmed that the distributions could not be separated. For this reason, it is not possible to compare results on the "Computer" versus "Cockpit" datasets in terms of head motion, though the two conditions involve different experimental conditions. Nevertheless, both datasets can be used to evaluate the difference between performing the identification on the eye-in-head positions versus the eye-in-environment. This is the purpose on the next section.

6.5.3 *Eye-in-head versus Eye-in-environment*

Both datasets included the eye-in-head position sequence *and* the eye-in-environment sequence. We run the ESIA on both sequences generating two ESVs. They were named ESV_{head} and ESV_{env} respectively.

The average proportion of samples being tagged as blinks, saccades, fixations in both datasets was compared (see Figure 45c).The proportions are remarkably similar for both datasets, around 85% of time spent fixating against 15% of time saccading, with 1-2% time spent blinking. Beck and Luck (2009) observed a similar distribution in the context of target detection on a computer screen: they found an average fixation rate of 4 fixations per second, with an average fixation duration of 200 ms. So fixations occupied, on average, 800ms per second of experiment, which corresponds to 80% of the total time while leaving 20% percent left for saccades and blinks.

In order to account for the differences between both ESVs in greater detail, frequency tables were established for both datasets (see Figure 45). The values are very close for both conditions ESV_{head} and ESV_{env}, in both datasets. The only notable difference is about the type of saccades. The "Computer" dataset involves comparable amount of small saccades and large saccades, whereas the "Cockpit" dataset involves no small saccades at all but only large saccades. In order to confirm the high correlation between both ESVs, we ran a statistical test. ESV contains discrete values, therefore we selected the Kendall Tau Correlation coefficient. Both tests were significant, with correlation coefficients close to 1. Results are presented in Table 6. These statistics confirms that both ESVs are highly dependent. We can therefore consider that it is equivalent to perform the ESIA identification using the eye-in-head sequence or the eye-in-environment sequence, in both experimental conditions.



(a) Bar chart representation

	"Computer" dataset				"Cockpit" dataset			
	ESV_{head}		ESV_{env}		ESV_{head}		ESV_{env}	
	#	%	#	%	#	%	#	%
0	817	1.0	817	1.0	1126	2.0	1128	2.0
1	4491	5.5	4277	5.2	9	0.0	11	0.0
2	5592	6.8	5427	6.6	8114	14.4	7903	14.0
3	2316	2.8	2074	2.5	1779	3.2	2005	3.6
4	13862	16.9	13329	16.2	9847	17.5	8779	15.6
5	6186	7.5	6023	7.3	6663	11.8	6135	10.9
6	48931	59.5	50248	61.1	28828	51.1	30405	53.9

(b) Frequency table for all 7 ESIA states. Percentage values were rounded but sum to 100%

	"Computer" dataset	"Cockpit" dataset
Blinks (0)	0.99	2.00
Saccades (1, 2)	12.04	14.23
Fixations (3, 4, 5, 6)	86.97	83.77

(c) Summary of frequencies. The two datasets were grouped, as the fixation and saccade subtypes.

Figure 45: Frequency table and associated bar chart representation for comparing ESV_{head} and ESV_{env}, in the "Computer" and "Cockpit" dataset.

Dataset	Kendall τ	p
"Computer"	0.9615	< 0.05
"Cockpit"	0.9401	< 0.05

Table 6: Kendall τ rank correlation table. Coefficient show an almost perfect correlation

6.6 ESTIMATION OF INTERFACE-DEPENDENT METRICS FROM THE ESIA

The previous section advocates for using the eye-in-head sequence as inputs of the ESIA. But discarding the eye-in-environment sequence prevents interface-related metrics such as the SWR to be computed. In this section we demonstrate that to some extent, the ESIA allows for an estimation of interface-dependent metrics such as the SWR.

The complete analysis of the ocular data relative to the environment was performed: each sample was therefore associated with one AOI (see Chapter 10-). Transitions from an AOI to another involves larger movements than small adaptations inside the same AOI. Therefore it seems natural that saccades associated with AOI transitions should be of greater extent, and with greater peak velocity than the average saccades. In order to test such an hypothesis, we isolated the saccades corresponding to the AOI transitions and compared their peak velocity to the peak velocity of all the saccades performed during the experiment. The results are plotted in Figure 46. The main mode of the velocity distribution of all saccades is around zero, whereas the main mode of the velocity distribution of the transition saccades is around $400°\,s^{-1}$. Most of the saccades with peak velocities above $300°\,s^{-1}$are transition saccades, whereas transition saccades represent only a small share of all saccades with velocities under $200°\,s^{-1}$.

As presented in Section 6.1.3, we proposed a $250°\,s^{-1}$threshold for separating small and large saccades. According to the velocity distribution, large saccades are likely to be transition saccades. There should be a correlation between the rate of large saccades detected by the ESIA and the SWR. To be validated, this hypothesis required the rate of large saccades to be computed to be comparable to the SWR. AOIs were robustly defined on the "Computer" dataset that was used in the first contribution presented in Chapter 5.

The switching rate between AOIs (SWR) was computed for every second of the experiment: the number of transitions from one AOI to another in the last 10.5 seconds was counted and divided by the duration of the POI, resulting in the SWR (expressed in number of shifts per minutes). Using the ESIA log, we computed the number of large saccades on the last 10.5 seconds for each second of experiment (same as for SWR), and compared to the SWR. Over the whole

dependency/histVsacVtrans.png

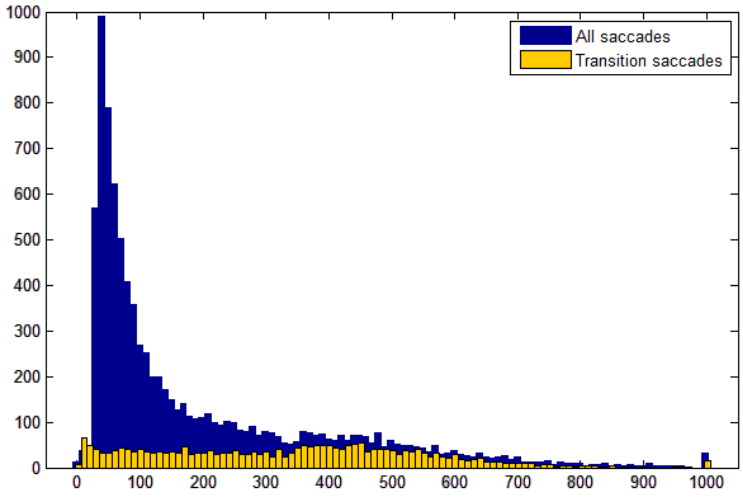


Figure 46: Histogram of velocities of transition saccades compared to all saccades. X axis in $^{\circ}s^{-1}$, Y axis in number of counts

dataset, a significant correlation was found, with a Spearman's rank correlation coefficient of 0.663.

6.7 CONCLUSION

The ESIA was defined explicitly in this chapter together with its derived metrics. The state distribution is a new metric that provides the contribution of each state in the ocular activity. Contributions are correlated, and two modes were elicited from the correlation map. These modes are consistent with the focal versus ambient model. Three ratios were defined afterwards. The ESIA was also tested for its ability to get the equivalent results by analysing the eye-in-head sequence versus the eye-in-environment sequence. Finally we showed the possibility to estimate interface dependent metrics such as the SWR from the ESIA. The next sections present use cases in which the ESIA and the derived ratios are calculated and used for detecting attentional impairments.

7

USE CASE 1: DETECTION OF ATTENTIONAL TUNNELLING

RÉSUMÉ EN FRANÇAIS

Le chapitre précédent à permis de valider théoriquement l'ESIA, il s'agit maintenant d'illustrer l'intérêt des métriques directes qui en découlent pour l'estimation des états attentionnels. Ce chapitre présente trois cas pratiques d'estimation d'état attentionnels en temps réel simulé, sans s'appuyer sur la connaissance de l'interface.

Les premier cas consiste en la détection de la tunnélisation attentionnelle au cours de l'expérience robotique du Chapter 5. Les principaux éléments de l'expérience sont rappelés succinctement en guise d'introduction, le détail de l'expérience ayant déjà été présenté. De même, le calcul de l'ESIA selon sa définition au chapitre précédent est considéré acquis. A partir des données de l'ESIA, il a été possible de suivre deux approches différentes pour l'estimation des états attentionnels par les métriques directes: l'approche focal versus ambiant, et l'approche exploitation versus exploration qui ont été présentées au Chapter 4.

L'approche focal versus ambiant classique consiste en l'analyse des couples fixation-saccade, mais des travaux plus récents suggèrent de considérer les couples saccade-fixation. Dans les deux cas, il est attendu que les fixations longues soient associées aux saccades de faibles amplitudes, et inversement. Les deux approches ont été considérées dans le cadre de cette expérience, 15828 couples fixation-saccade et 15953 couples saccade-fixation ont été identifiés sur l'ensemb-le de l'expérience. Dans un premier temps, la distribution des couples est étudiée au travers d'un algorithme de k-clustering, en prenant la valeur 2 pour le coefficient k. Les deux clusters mis en évidence ont une valeur moyenne de durée de fixation similaire (autour de 260ms) alors que les amplitudes moyenne des saccades diffèrent (30° pour 15Dans un second temps, les trois ratios exploitation/exploration sont calculés d'après leurs définitions données au chapitre précédent. Pour chaque participant, le ratio s'appuie sur la distribution des états oculaires calculée grâce à l'ESIA. Conformément à l'analyse de corrélation des états oculaires qui est a été faite au chapitre précédent, les fixations longues s'effectuent au détriment des saccades, et inversement. Les ratios amplifient cette tendance par un effet multiplicateur lié à leur définition. On remarque par exemple que les phases d'exploitation des informations du bandeau vidéo correspondent aux valeurs les plus élevées du ratio, alors

que les phases de recherche d'information sur l'interface correspondent aux valeurs les plus faibles (notamment lors de changement de tâche au cours de la mission). Il est finalement vérifié statistiquement que le ratio a bien une valeur significativement plus élevée pour le groupe TUN=1 que pour le groupe TUN=0, ce qui en fait une métrique pertinente de la tunnélisation attentionnelle. Ce ratio a par ailleurs été utilisé dans une étude du BEA. Onze équipages professionnels (pilotes PF et co-pilote PNF) ont effectué un vol de deux heures et demi en simulateur complet de classe D (A330 ou B777) comprenant 3 remises des gaz. Les activités oculaires du PF et du PNF étaient mesurée simultanément à 50Hz via deux oculomètres mobiles synchronisés. Douze aires d'intérêt (AOI) ont été définies couvrant les principaux instruments de bord et le pare-brise. Indépendamment de cette approche "AOI-based", l'ESIA a été utilisé pour calculer les métriques "directes" classiques et le ratio exploration versus exploitation normalisé en z-score sans aucune prise en compte de l'interface. Empiriquement, il est remarqué que lorsque que le ratio prend des valeurs supérieures à deux écarts types, le pilote ou co-pilote a un comportement dont les caractéristiques sont similaires à la tunnélisation attentionnelle pendant plus de 10 secondes. En appliquant ce critère à l'ensemble des enregistrements, il a été possible de cibler rapidement ces moments d'intérêt de quelques secondes dans plusieurs heures d'enregistrement.

ENGLISH VERSION

In chapter 4, we implemented machine learning techniques and AOI based metrics to detect attentional tunnelling. Though our results were convincing, this approach is limited as long as our metrics are interface/tasks dependent. We therefore proposed the implementation the ESIA in the last chapters, an algorithm that computes 7 eye states. This approach is tested with the data collected during the same experiment involving the robot battery failure (see Chapter 5). The seven "eye states" and other derived metrics such as ambiant/focal and fixation/saccades ratio are benchmarked. The fixation saccades ratio is shown to be a relevant indicator of attentional tunneling.

7.1 INTRODUCTION

This section aims at reminding the main elements describing the experimental set-up, participants, and grouping. All these elements were detailed in the Chapter 5 of this thesis.

7.1.1 *Scenario*

The mission was separated in four segments. During the first two segments, the robot progressed autonomously to reach its destination. Once the destination reached (start of the third segment), the participant took hand over the robot. Participants were tasked to identify a visual target using the robot panoramic video. During this demanding task, a battery failure was triggered (start of the fourth segment). Participants were expected to abort the manual operation of the robot and let the robot go back to base autonomously. After 50 seconds, the mission was stopped (end of the fourth segment). Depending on the participants' decision, either the robot was on its way to the base (mission success), or the participant was still handling the robot for the target identification despite the battery alarm (mission failure).

7.1.2 *Participants, behavioural result and grouping*

23 participants from our campus were recruited for this experiment. 12 participants were assigned to the control group, and 11 to the countermeasure group. In the control group, participants had no dedicated help to detect the battery failure, whereas participants in the countermeasure group benefited from one. Among the 12 participants from the control group, 8 did not notify the battery alarm for 50 seconds. They were assigned to the TUN=1 group. Among the 11 participants from the countermeasure group, all detected the alarm but one did not respect the procedure and kept on piloting the robot until the battery run out. His data were not considered for this exper-

iment. In this use case, the focus is on comparing participants from the TUN = 1 group versus participants from the TUN = 0 group.

7.1.3 *Measurement and metrics computation*

All participants were equipped with the EyeTechSensor from Pertech, recording at 25Hz with a precision of 0.25° (manufacturer information). Though AOIs were defined in this experiment, we consider only the eye-in-head positions streamed by the eye-tracker. This sequence was produced in real-time (after the eye-tracker calibration), and stored in a log. All analyses were computed off-line, but all the metrics that are presented in this use case could be computed on line with an appropriate implementation of the ESIA. In our case, the ESIA ran on a Matlab environment, but on-line analysis was simulated by considering samples one after the other, as it would be the case in a real on-line process.

7.2 FOCAL VERSUS AMBIENT

As presented in the previous section, focal versus ambient fixation classifications are supposed to be relevant indicators of the attentional process. Focal denotes the in-depth encoding of visual information for high-level ventral processing, while ambient denotes the fast spread-out dorsal encoding of the visual scene. This classification was proven to be especially relevant in the context of free-viewing of natural scene (Unema et al., 2005, Follet et al., 2011, Velichkovsky et al., 2005), but also in the context a dynamic scene perception (Velichkovsky et al., 2002). We wanted to test wether this classification could apply in the context of a robotic experiment.

7.2.1 *Computation of the fixation-saccade couples*

The study of focal versus ambient fixations requires to group one fixation and one adjacent saccade to form couples. These couples are then used for eliciting relations between saccade extent and fixation duration. Though Unema et al. (2005) recommended to form the couples from fixations and their subsequent saccades, Follet et al. (2011) suggested the opposite.

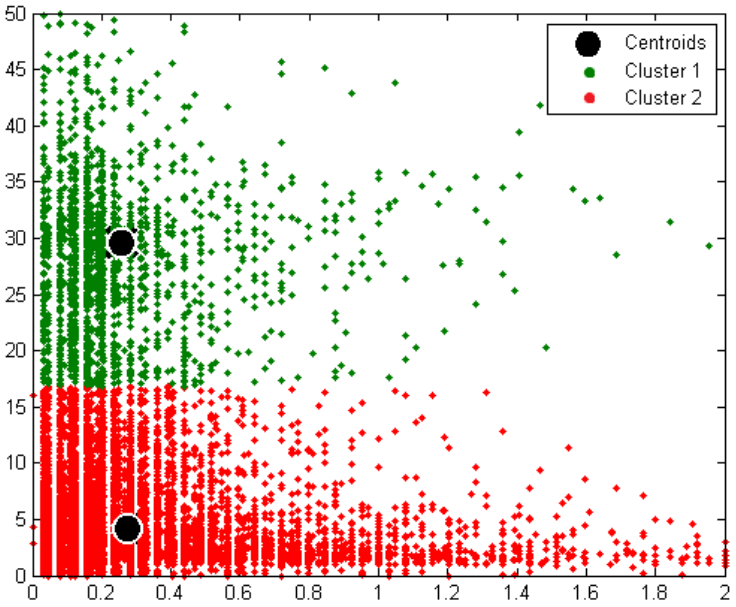
Both configuration were considered and compared for forming the couples. From the the ESIA log computed on all participants from both groups, we formed couples or fixation-saccade, or saccade-fixation. For forming saccade-fixation couples, we used the event type field in the log for isolating the saccades (coded 1 or 2), and tested the subsequent event type. If the event type code was 3, 4, 5 or 6 (code for all type of fixations) then a couple was formed as the concatenation of the two log lines. In total, 15828 fixation-saccade couples and 15953

saccade-fixation couples were formed. The difference comes from the presence of blinks that interfere with the pairing process.

7.2.2 *Saccade extent as a function of fixation duration*

We followed a similar approach to (Follet et al., 2011). From the fixation-saccades couples, we elicited two clusters using k-means clustering ($k = 2$). Results are presented in Figure 47. As in their study, the two clusters show similar fixation durations, but different saccade extents, which advocates for using the saccade extent as the indicator for classifying couples as focal or ambient. Centroids of the clusters present differences with the clusters found by (Follet et al., 2011). The relative sizes of both clusters (15% cluster 1 versus 85% cluster 2) are also different from the ratio found in their study (30% cluster 1 versus 70% cluster 2). Both differences could be due to the difference in experimental conditions between both studies (free-viewing of natural scene against robotic supervision). In this clustering, cluster 1 should be associated with ambient fixations, whereas cluster 2 should be associated with focal fixations.

In order to further investigate this possible clustering of the population, we considered plotting the extent of the next and previous saccades of each fixation as a function of the fixation duration. We also considered plotting the "inverse" profile, as our clustering suggests that saccade extent should be the indicator for segregating focal and ambient fixations. This profile consists of the fixation duration being plotted as a function of the next or previous saccade. Both profiles are presented in Figure 48. According to the reference work on that field (Unema et al., 2005, Follet et al., 2011, Velichkovsky et al., 2005), we expect the saccade extent to be of greater extent (10-20 °) for fixation of shorter duration (less than 200-250 ms). Also, though the "inverse" profile was not proposed in these studies, we expect the fixation duration to be longer with smaller saccades. First, there is almost no effect of using the previous versus the next saccade for forming the fixation-saccade couples (solid and split lines are superimposed). Second, there is almost no effect of the fixation duration nor of the saccade extent, the profiles are flat which does not comply with our expectancies. Third, the standard deviation is of greater extent than the profile that can be found in the reference studies. These results suggest that the focal versus ambient pattern was not dominant in our experiment compared to other patterns. In other words, in our experiment, the duration of the fixation does not predict the extent of the saccade, and vice-versa. We therefore did not investigate this approach further, but investigated the state distribution instead.

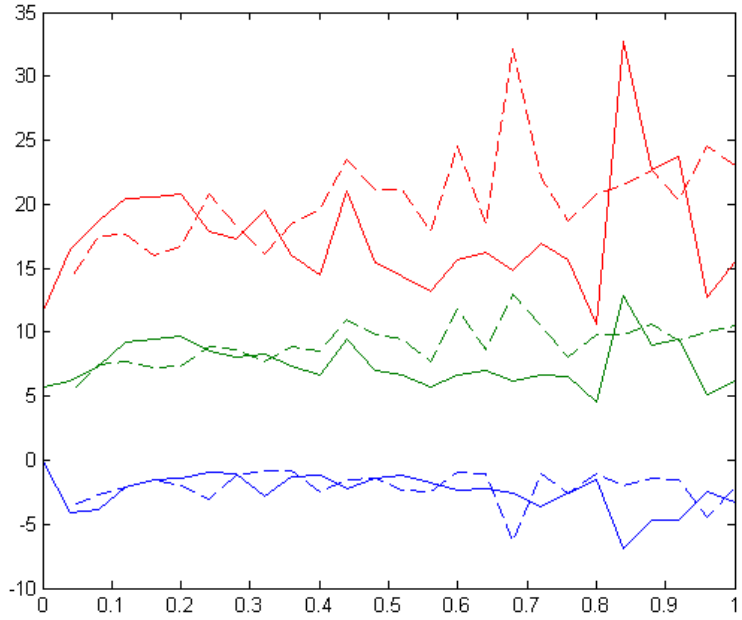


(a) Scatter plot of saccade extent (Y axis) depending on the previous fixation duration (X axis).

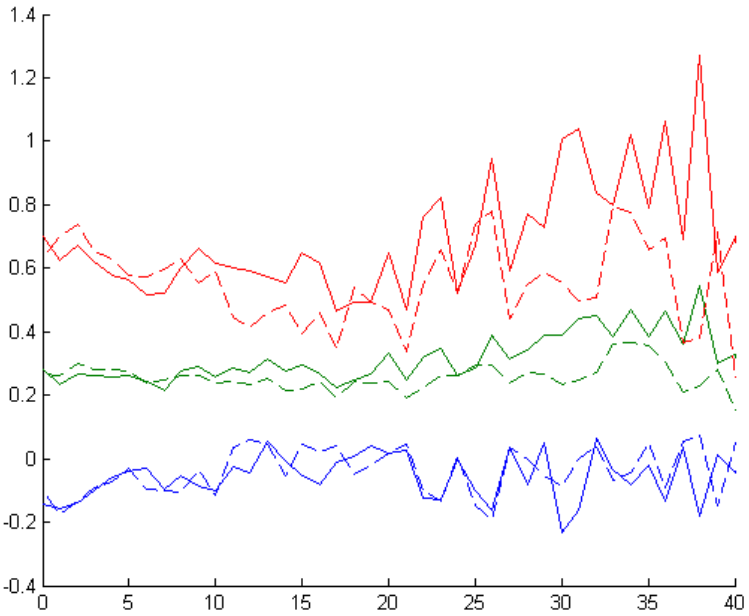
Cluster	Fix. Dur. (ms)	Sac. Ext. (°)	# of cases	% of cases
1	255,5	29,61	2368	14,84
2	273.1	4,14	13,585	85,16

(b) Clusters characteristics

Figure 47: Scatter plot of the saccade extent (Y axis) as a function off the previous fixation duration (X axis). Total population of 15828 fixation-saccade couples. Two clusters were elicited using k-means clustering with $k = 2$. Clusters characteristics are in the table.



(a) Saccade extent (Y axis, in degrees) as a function of fixation duration (X axis, in milliseconds).



(b) Fixation duration (X axis, in milliseconds) as a function of saccade extent (Y axis, in degrees).

Figure 48: "Focal versus ambient" profiles. These relations were established for both the previous saccade (split lines) and the subsequent saccade (solid lines). The central green lines are the mean, whereas red and blue lines are the envelopes defined as the mean +/- the standard deviation.

7.3 STATE DISTRIBUTION FROM THE ESIA, AND RATIOS

7.3.1 *Comparison of state distribution*

One interesting metric from the ESIA is the state distribution. We computed this state distribution for all participants. In order to get a meaning of the difference between participants from the $TUN = 0$ and the $TUN = 1$ groups, a comparison of two example of distributions can be found in Figure 49.

7.3.2 *Ratio Calculation*

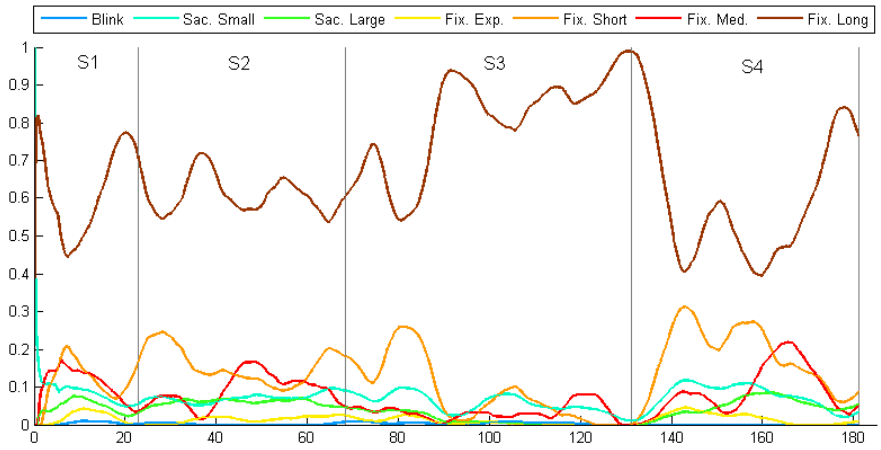
The ratios defined in the previous section were calculated for all participants. Ratio values for the two participants taken as examples in Figure 49 are plotted in Figure 50

Finally, the informative power of the state distribution seems to be kept with this aggregated metric, which can be used as a summary. The exploration and exploitation phases highlighted are consistent with the participant behaviour on the different phases of the mission. In order to confirm this supposed link between the ratio and the attentional state of the operator, we conducted a two-way analysis of variance.

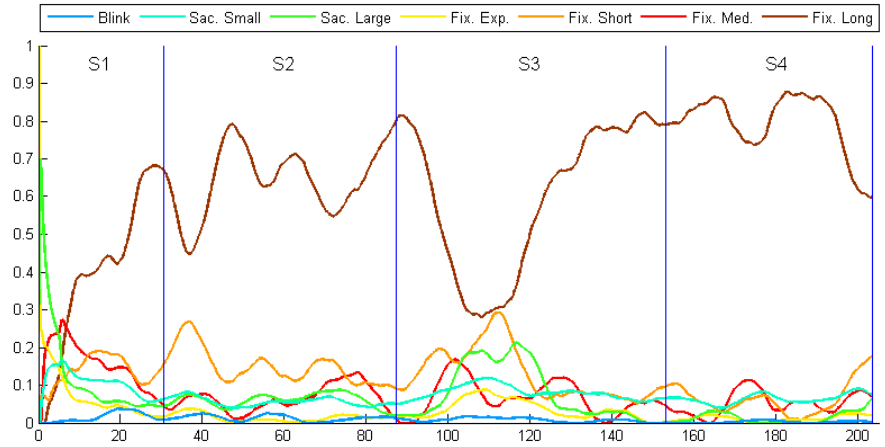
7.3.3 *ANOVA of the exploit/explore ratio*

Groups were used as an inter-participant categorization variable, whereas segments were used as an intra-individual categorization variable, with the ratios being the dependent variables. Effect of the Segment * Group interaction was found significant for the $R_{N-correl}$ and R_{N-ESIA} but not for R_{N-gold} (results are presented in Figure 51). We considered a Tukey's Honest Significant Difference (HSD) post-hoc analysis for pairwise comparison. We selected this test against the Fischer's Least Significant Difference (LSD) for its stricter corrections on the p-values. We focused our analysis on comparing the differences between groups on each of the 4 segments. Results are presented in Table 7. The first two segments show no difference between group as seen on Figure 51. The differences observed between groups at segment 3 are not significant. Greatest differences are observed in S4 which was expected to be the differentiating segment, but R_{N-gold} and $R_{N-correl}$ are not significantly different under the HSD test. Only the R_{N-ESIA} is significantly different between the two groups.

In other words, the R_{N-ESIA} is the only ratio out of the three that is statistically able to differentiate a participant from the $TUN = 1$ and a participant from the $TUN = 0$ group. The difference between groups was the fact that participants faced attentional tunnelling or not. We

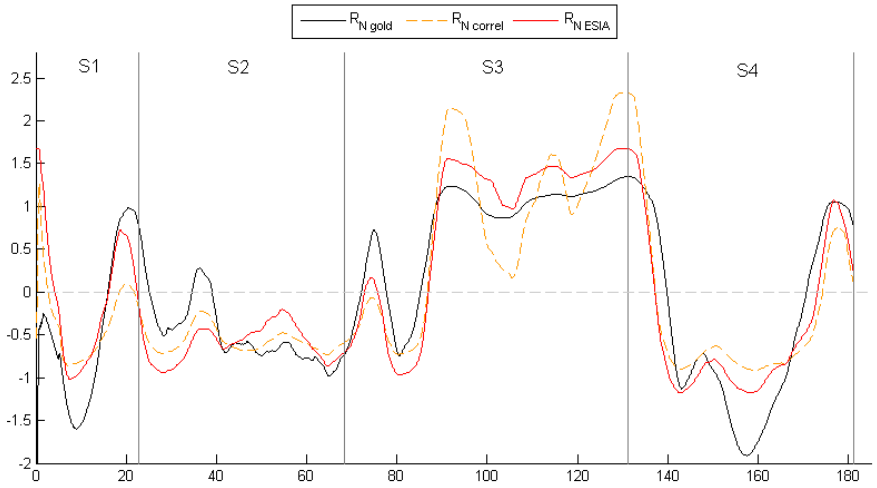


(a) Participant A from the TUN = 0 group. The participant had an assistance for detecting the battery alarm between segment 3 and 4, and detected the alarm immediately. The first two segments seem quite homogeneous. During the third segment, associated with the manual piloting of the robot, a change is noticeable: long fixations occupy the majority of the ocular resource, reaching close to 100% when the battery alarm pops out. Thanks to the assistance, the participant noticed the alarm immediately. This results in an active search of the right information to apply the procedure as shown by the ocular activity: both types of saccades and short fixations occupy more than the half of the ocular resource.

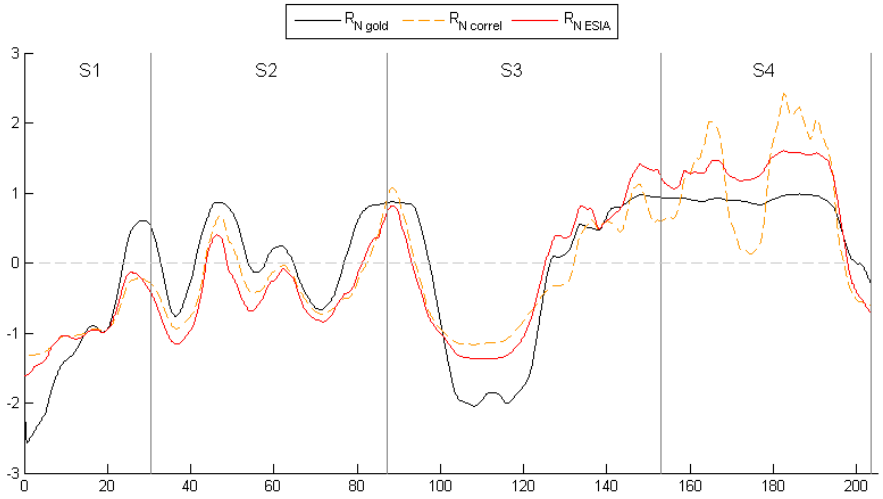


(b) Participant B from the TUN = 1 group. The participant had *no* assistance for detecting the battery alarm between segment 3 and 4, and kept on piloting for 50 seconds after the alarm. Three main differences can be noted. First, at the beginning of the experiment, values seem unrealistic with a very high saccade ratio. This is a computational artefact. The state distribution is calculated on a window of ten seconds, but at the beginning, there are not enough recordings for filling the 10 second window, resulting in an unreliable outcome. Second, a strong drop in long fixations is observed 10-20 seconds after the beginning of segment 3. It corresponds to the operator shifting from autonomous to manual, which led them to reconsider the whole interface in order to recall the main instruments and actually start the manual piloting. Third and last, contrary to the other participant, no shifts are observed when crossing from S3 to S4 (battery alarm). The operator indeed did not notice the battery failure, and kept the same strategy for identifying the target for 50 seconds until the robot ran out of battery.

Figure 49: Example of the temporal state distribution of two participants from the robot experiment. For each second of the experiment (X axis, in seconds), the state distribution (Y axis, in real values between 0 and 1) was calculated on a moving POI consisting in the last 10 seconds. Filtered with a 5 seconds moving average. Vertical lines represent the mission triggers between each segment of the mission.

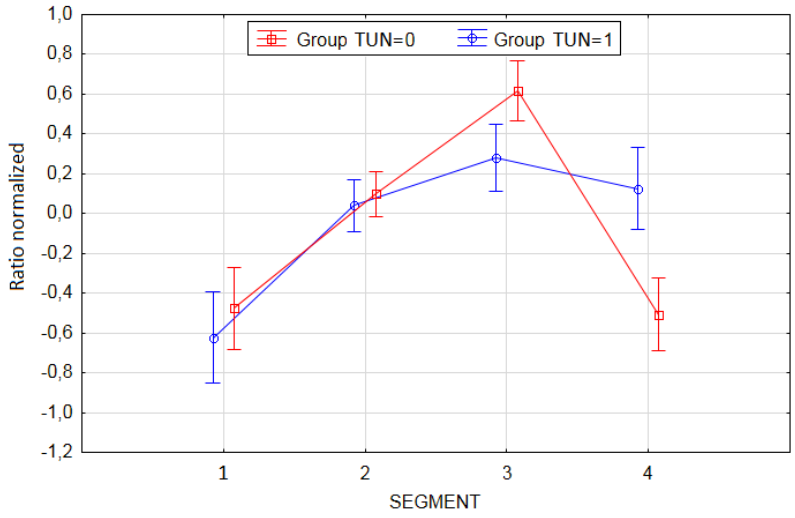


(a) Participant A from the TUN = 0 group. The participant had an assistance for detecting the battery alarm between segment 3 and 4, and detected the alarm immediately. During the third segment, which is associated with the manual piloting of the robot, a change is noticeable: the ratios show a strong exploitation of the visual information reaching its maximum when the battery alarm pops out. Thanks to the assistance, the participant noticed the alarm immediately. This results in a ratio value indicating exploration, that is associated with the diagnosis of the failure and the shift from manual piloting to passive monitoring.

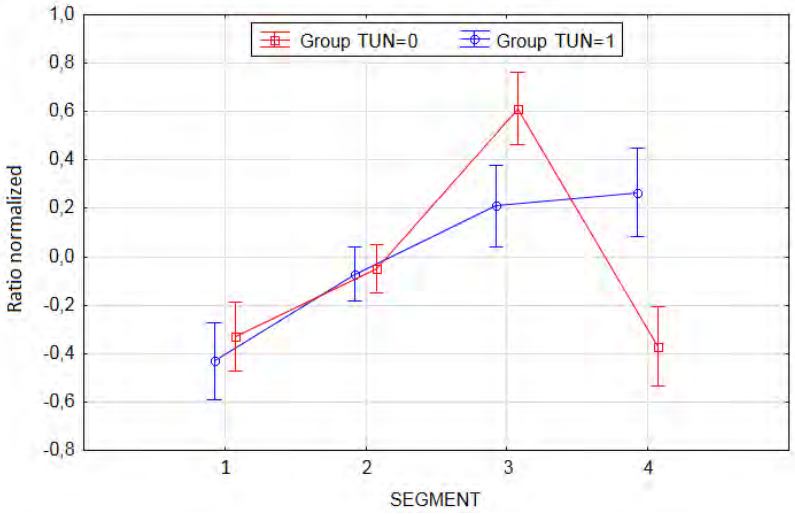


(b) Participant B from the TUN = 1 group. The participant had *no* assistance for detecting the battery alarm between segment 3 and 4, and kept on piloting for 50 seconds after the alarm. During segment 3, the ratios reflect exploration. This was identified as a period in which the subject shifted from autonomous to manual. When the battery alarm pops out, no change in the ratio is to be found. This is consistent with the fact that the operator did not notice the battery failure, and kept the same strategy for identifying the target. At the end of the record, a drop is also noticeable, but is not interpreted here.

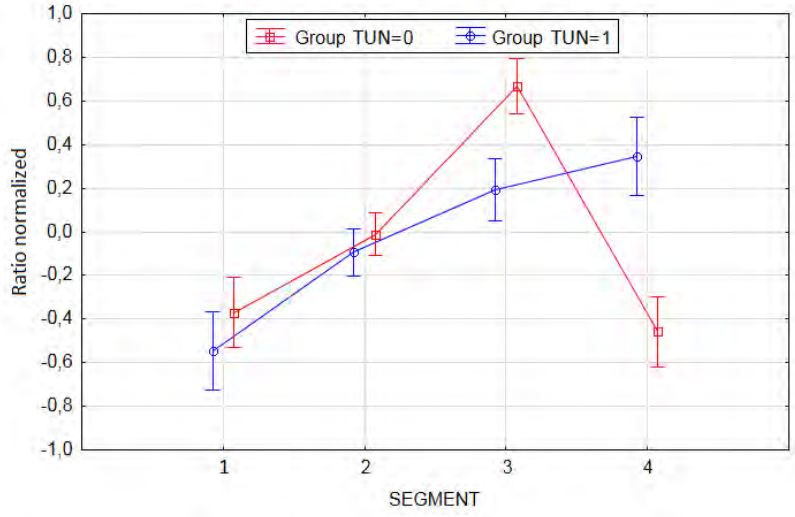
Figure 50: Example of the three exploit/explore ratios calculated for two participants of the robotic experiment. For each second of the experiment (X axis, in seconds), the ratios (Y axis, in real values) were calculated on a moving POI consisting in the last 10 seconds. Smoothed with a 5 seconds moving average. Vertical lines represent the mission triggers between each segment of the mission.



(a) R_{N-gold} SEGMENT*Group effect: $F(3,48) = 2.3240, p = .08670$



(b) $R_{N-correl}$ SEGMENT*Group effect: $F(3,48) = 3.6384, p = .01913$



(c) R_{N-ESIA} SEGMENT*Group effect: $F(3,48) = 5.8515, p = .00172$

Figure 51: ANOVA for the interaction Group*Segment on the three ratio. Unweighed means. Vertical bars denote +/- standard errors.

Ratio	S1	S2	S3	S4
R_{N-gold}	-	-	-	-
$R_{N-correl}$	1,000	1,000	0,551	0,070
R_{N-ESIA}	0,990	1,000	0,319	0,007

Table 7: Results of the Tuckey's HSD for comparing ratio values between the TUN = 1 and TUN = 0 groups on each segment of the robotic mission. Only the R_{N-ESIA} on S4 is significant. R_{N-gold} pairwise comparison was not conducted as the interaction was not significant (see Figure 51a)

can therefore conclude that the ratio is a relevant attentional metric able to identify the participants who faced attentional tunnelling on that mission, using only interface-independent metrics from the ESIA.

7.4 TEST OF THE RATIO IN A COMPREHENSIVE BEA STUDY

The French Safety Board (Bureau Enquete Analyses - BEA) conducted a simulation study on go-around manoeuvres (BEA, n.d.) in which the ESIA was used for computing the $R_{N-correl}$ ratio. This section presents a summary of the study and focus especially on the contribution of the ratio for identifying periods of special interest from an human factors point of view.

7.4.1 *Abstract from the report*

Events during which a safe go-around was not achieved have been brought to the attention of the BEA due to their noticeable increase since the 2000s. Considering the greater risk associated with the go-around manoeuvre, which is currently defined as a normal phase of flight, the BEA has decided to undertake a study named ASAGA (Aircraft State Awareness during Go Around). The study aimed at identifying the weaknesses of the system during go-around manoeuvres regardless of the type of aircraft and the airline operational procedures. In this context, the analysis relied mainly on three components: (1) Safety events (incidents / accidents); (2) Large scale survey; and (3) Flight simulations. Experiments were conducted using two long-range full flight simulators with operational pilots. This section presents the method used to conduct the experiment and some of the resulting data, as well as some first conclusions. The data included: video camera recordings for task analysis; recordings of the ocular activity of the two pilots; and debriefings. The eye trackers were used to analyze the pilots' scanning patterns and the distribution of their attention during this particular flight phase (see Figure 52). The de-

Figure 52: Inside the A330 full flight simulator during the BEA experiment.

briefings consisted of interviews based on the self-confrontation technique, which enable the pilots to self-assess their actions and situational awareness. Eleven crews from three French airlines participated in the simulations. The scope on the experiment was not explained to them to prevent any anticipation. Each crew performed the same scenario from take off to landing that included three go-around manoeuvres. The first results indicate some concerns related to the cooperation between controllers and aircrews in such demanding occurrences, some lack of understanding of cockpit automation and failures in crew resource management.

7.4.2 *Summary of the study*

The experiments were conducted using two long-range full flight simulators (B777 and A330). Three types of data were collected :

DIRECT OBSERVATIONS were made during the experiment and from video recordings: video (HD) was used to record all the displays and all the communications to allow an accurate post-analysis. Observations were carried on during the entire scenario by two human factors specialists inside the cockpit and by two investigators with the help of a remote video streaming.

SUBJECTIVE ASSESSMENT The debriefings consisted of interviews, based on the self-confrontation technique, which enabled the pilots to self-assess their actions and their situational awareness.

VISUAL SCAN The objective was to analyze the pilots' scanning patterns and the distribution of their attention during the go-around. Two synchronised 50Hz Pertech head-mounted eye trackers were used (0.25 degree of accuracy). 12 areas of interest (AOIs) on the cockpits were defined as follows: 1) primary flight display, 2) speed, 3) artificial horizon, 4) altitude, 5) heading, 6) flight mode annunciator (FMA), 7) navigation display (ND), 8) electronic centralized aircraft monitor (ECAM), 9) left multipurpose

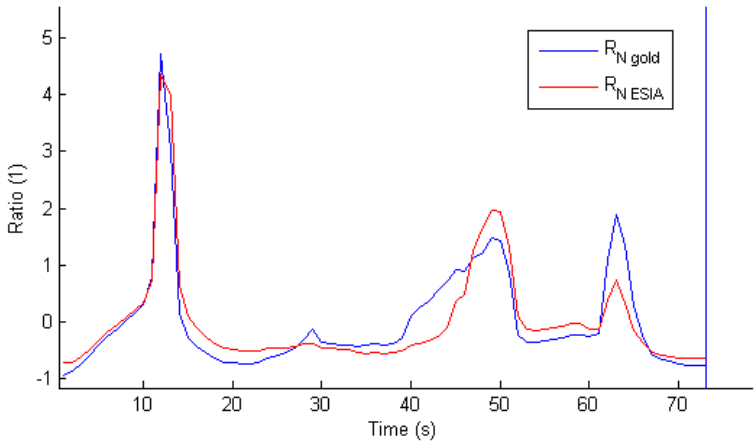


Figure 53: Ratio calculation from the Crew 10's eye-tracking data. The ratio was calculated on a window of 10 seconds. The first peak corresponds to the PF being concentrated exclusively on the flight direction for 11 seconds.

control display unit (MCDU), 10) right MCDU, 11) visual scene, 12) flight control unit (FCU). A 13th AOI was considered in order to collect the ocular fixations out of the other twelve previous (out of zone, OZ).

All participants were volunteer. Each session lasted at least 4h30 and consisted of the following steps: briefing (30 min); installation in the cockpit (equipped with the data recording kits) (30min); flight (2h); debriefing (1h30).

7.4.3 *Use of the Ratio*

The ESIA was run on all ocular data collected in this experiment. Only the $R_{N-correl}$ and $R_{N-correl}$ were computed on this experiment, and we detected that for rare and shorts periods, the ratio was taking values above 2 (i. e. 2 standard deviations). An example of such measure on a go-around is presented in Figure 53. These periods were isolated and the associated flight records were studied thoroughly.

7.4.4 *Corresponding situations in the cockpit*

According to the human factors experts in the team, these periods indeed corresponded to hazardous behaviours during the go-arounds. Especially four events rose a specific interest and were explicitly reported in the final report, i. e.:

Crew 1, first go-around. During this go around, the PNF/PM set a heading of 240° instead of 340°into the MCP.

> The PF tracked the flight director to the incorrect heading for 11 s until the controller reminded the crew of the correct heading to advise them of their error. The PF realised immediately that he should have flown the aeroplane in the lateral plane directly to the heading of 340, without using the flight director ("I should have flown to the heading myself");
>
> *Crew 7, first go-around.* During this go around, the PF focussed for 22 s on the vertical flight path to capture the target altitude, to the detriment of his monitoring of the horizontal flight path from which the aeroplane was deviating. Indeed, analysis of the eye-tracking data revealed that the PF's gaze switched exclusively between the altimeter, the pitch bar on the flight director and the FMA.
>
> *Crew 9, first go-around.* During this go around, the PF took responsibility for setting the altitude and heading values on the FCU. As a consequence of focusing on this instrument he failed to see, on his own PFD, that the aeroplane was significantly overshooting the altitude specified (i.e. 2,500 feet). The indicators of the PF displaying some level of attentional tunnelling are his inability to anticipate this altitude overshoot, despite having fixated on the altimeter for a long time before entering data into the FCU, and despite his co-pilot calling out this deviation ("We're climbing through 3,000 feet").
>
> *Crew 10, first go-around.* During this go around, the PF concentrated exclusively on the flight director for 11 seconds. He never looked at the FMA, did not notice that "Heading" mode had been engaged by the PNF, and was not aware of the mode transition to "Navigation". As a result, he tracked the flight director in the lateral plane, despite being in "Navigation" mode (heading 330) or in "Heading" mode (heading 340), even though he had called out that the heading was 340.

This demonstrated another possible use of the ratio in this study, which is to identify *short periods of few seconds* of dangerous over focalization of the pilots visual attention *in long recordings of 2 hours*. To our knowledge, this is the first analysis of this kind. In this experiment, we isolated empirically the periods corresponding to ratio values above 2 standard deviations. This approach allowed to focus on these events first, and saved lot of time that are usually dedicated to the linear analysis of the video recordings (around 50 hours of recording in total).

7.5 CONCLUSION

In this chapter, we tested the ESIA metrics on a realistic robotic experiment. Two groups of participant were isolated, one that faced attentional tunneling and one that did not. We first considered the ESIA log for forming fixation-saccade and saccade-fixation couples. We used a similar k-means clustering as in (Follet et al., 2011) in order to form polls of fixations. Differences between polls were found on the saccade extent. In order to further investigate such differences, we plotted the profiles of the saccade extent versus the fixation duration and vice-versa. These profiles did not elicit the classical relationship that are expected when looking for the focal versus ambient classification. We diverted our study to the ESIA state distribution. This provided interesting ocular metrics that were consistent with the participant's behaviours. A correlation analysis showed that long fixations were negatively correlated to all other states, whereas saccades and short fixations were positively correlated. This suggested grouping led us to consider again the two modes of processing and make a parallel with the exploit/explore ratio that was defined by Goldberg. Thanks to the correlation analysis and to the temporal classification proposed with the ESIA, we could suggest 3 definitions of the exploit/explore ratio. These 3 ratios were computed on all participant's data, and compared across groups and segments of the robotic experiment. Finally, only the ratio defined as suggested by the temporal classification of the ESIA was significantly different on the fourth segment of the mission, as it could identify participants who faced attentional tunnelling against participants who did not. The BEA study was a realistic successful test of the usefulness of the ratio. It identified anomaly in the ocular activity in the long records, and allowed saving time to identify efficiently dangerous behaviours in the cockpit. These periods we further analysed by the BEA's human factors experts and indeed corresponded to risky situations.

8

USE CASE 2: DETECTION OF ATTENTIONAL DISORIENTATION

RÉSUMÉ EN FRANÇAIS

Le cas précédent a démontré l'intérêt du ratio exploitation versus exploration pour la détection de la tunnélisation attentionnelle: les valeurs anormalement élevées du ratio traduisent en effet que l'exploitation excessive d'une infor-mation se fait au détriment de l'exploration des autres sources potentiellement pertinentes. Dans ce chapitre, il est montré que le ratio est également sensible à la désorientation attentionnelle et permet donc sa détection. Celle-ci peut-être considérée comme "l'opposé" de la tunnélisation attentionnelle: elle correspond à une recherche excessive car inefficace des infor-mations pertinentes au détriment de leur exploitation. Dans cette simulation, seize participants ont effectué un vol au pilote automatique en suivant les instructions du contrôleur aérien. Au cours du vol, trois transitions "cachées" de mode du pilote automatique qui vont successivement provoquer un "conflit" avec les intentions du pilote. Les pilotes, lorsqu'ils détectent que l'avion ne suit pas le comportement attendu, essayent de diagnostiquer l'origine du problème. La phase de diagnostique s'accompagne de la recherche des infor-mations pertinentes sur les instrument de bord. Le ratio prend alors des valeurs faibles qui traduisent cette recherche d'infor-mation. Un des trois conflit a néanmoins provoqué la désorientation attentionnelle de plusieurs pilotes qui n'ont pas su diagnostiquer la panne pendant plusieurs secondes. Il est finalement montré statistiquement que ces phases de désorientation attentionnelle sont associées à des valeurs de ratio plus faibles de 5 écarts type. Le ratio permet donc la détection de la désorientation attentionnelle, en suivant une approche similaire à celle présentée au cas précédent avec la tunnélisation attentionnelle.

Finalement, l'ESIA permet le calcul de métriques de l'état attentionnel des opérateurs en mission quelques soit le contexte: à partir d'un eye tracker mobile générique et en utilisant les mêmes paramètres pour l'algorithme dans des domaines d'application différents, on remarque que le ratio garde son pouvoir informatif, notamment pour la détection de la tunnélisation attentionnelle et la désorientation attentionnelle. Par ailleurs, l'ESIA permet un gain de temps remarquable tant du point de vue du traitement des données oculaires que du point de vue de l'analyse des enregistrements de longue durée.

ENGLISH VERSION

In the previous use case, we have shown that the ESIA could identify attentional tunnelling. In this study, we intent to reveal that the ESIA could detect another impaired attentional state that we named "Attentional disorientation". This attentional state is the "opposite" of attentional tunnelling and is characterized by an excessive and inefficient visual search. This demonstration was made during an experiment in ISAE flight simulator with pilots facing a critical conflict with automation. This work was submitted at CHI2015 conference.

8.1 MATERIAL AND METHODS

8.1.1 *Participants*

Sixteen healthy volunteers (two females; mean age = 30.8, SD = 14; mean flight experience = 1391 hours, range 55–9000; mean automated flight desk experience 200 hours, range 15–700), all French defence staff from French Aerospace engineering School (ISAE) participated in this experiment after giving written informed consent. Participants had good auditory acuity and normal or corrected-to-normal vision.

8.1.2 *Flight simulator*

The experiment was conducted with ISAE 3-axis motion flight simulator (see figure 54). The user interface was composed of the classical electronic flight instrument displays including the primary flight display (PFD) and the navigation display. The participants had a sidestick, rudder pedals, thrust levers and a Flight Control Unit (FCU) to control the flight guidance. The FCU was dedicated to interact with the autopilot thanks to four knobs (speed, heading, altitude, vertical speed). The autopilot (vertical and horizontal profile management) was engaged and disconnected via a push button "AP" on the FCU. Autopilot disconnection was associated with a dedicated auditory warning ("cavalry charge"). The auto-thrust (speed management) was engaged and disengaged via a push button "ATHR" on the FCU.

For the purpose of the experiment, we implemented an autopilot which behavior was inspired from different modern autopilot systems (Pizziol et al., 2014).

- The autopilot had one lateral mode ("Heading") and three vertical modes ("positive vertical speed", "negative vertical speed", "null vertical speed"). Note that the vertical speed was null (i.e. 0 $ft.mn^{-1}$) when the autopilot reached the target altitude or when the pilot pushed the vertical speed knob to level off. These dif-

Figure 54: ISAE 3-axis motion flight simulator.

ferent flight modes were displayed on the upper part of the PFD;

- The autopilot automatically disconnected during over-speed or low speed/stall events. These events triggered the speed auditory warning (triple chime) that inhibited the autopilot disconnection auditory warning;
- The aircraft leveled off in case of inconsistent programming of the vertical speed with regards to the target altitude (i.e. it was not possible to "climb" with a negative vertical speed or to descend with a positive vertical speed).
- the autopilot reversed from level off or negative vertical speed to positive vertical speed when the aircraft speed was 5 knots close to the maximum speed (Vmax). This mode reversion was dedicated to anticipate a possible over-speed.

8.1.3 *Experimental scenario*

The scenario, that included the occurrence of one conflict with automation, lasted 10 minutes. The ATC (Air Traffic Control) cleared the pilots for take-off from Blagnac airport (Toulouse, France) and the airplane was vectored regularly according to a flight plan that was identical for each participants. At 9000 ft, the ATC required the participant to "accelerate 325 knots, descend 5000 feet with a -1000 $ft.mn^{-1}$ vertical speed" to avoid an incoming aircraft. As the speed reached 325 knots (i.e. 5 knots below maximum cruise speed), the autopilot reversed to positive vertical speed mode (+1000 $ft.mn^{-1}$) to

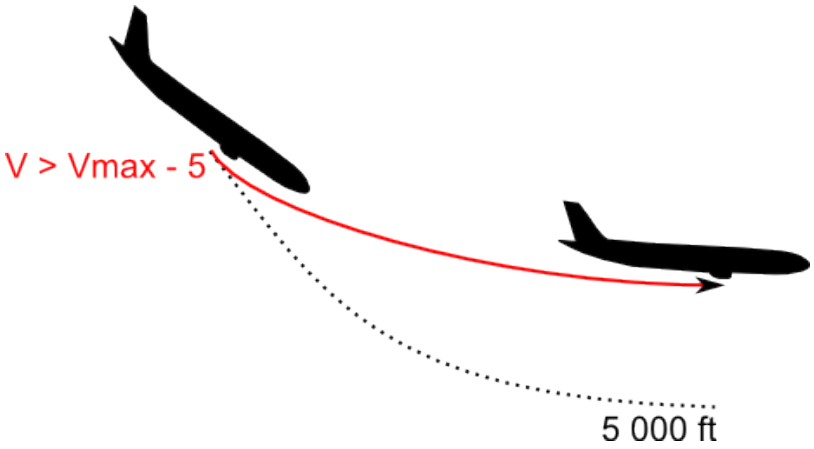


Figure 55: [The "impossible descent" automation surprise.] The "impossible descent" automation surprise: the ATC required the participants to descend but with an excessive speed. The combination of two autopilot behaviours led the aircraft to level off to prevent over-speed.

anticipate potential over-speed. This situation led eventually the airplane to level off instead of descending as the selected target altitude (i.e. 5000 feet) could not be reached with a positive vertical speed (see figure 55).

8.1.4 *Procedure*

The participants had a 20-minute tutorial presentation that detailed the flight simulator user interfaces (PFD, FCU, etc.) with a special focus on the different nominal and off-nominal autopilot behaviors. In order to check the understanding of the tutorial, the participants had to comment each slide of the tutorial and to recall the autopilot behavior. The volunteers completed a one-hour training in the flight simulator that included basic manual flying, landings, take-offs, stall and overspeed recovery. They were then trained to interact with the autopilot and had to set different parameters according to ATC instructions (e.g. "Supaero32, steer 200 degrees, climb 6000 feet, vertical speed +1000 $ft.mn^{-1}$"). Eventually off nominal situations such as autopilot automatic disconnection due to over-speed/stall events, inconsistent FCU programming, level off to vertical speed mode reversion were induced and the participants were told to recover from them. After the training, the participants had to repeat once the way to interact with the FCU and the autopilot behaviour as well as the different auditory warnings. The experimental scenario was then about to start.

8.1.5 *Behavioural Data Analysis*

Flight parameters, lateral and vertical modes of the autopilot and pilot's action on the FCU knobs were collected. The volunteers were also debriefed right after the end of the experiment and were told to explain their understanding of the autopilot behaviour during the descent and to comment their actions. The analysis of these data allowed us to characterize conflict occurrence, and conflict solving. As stated in the introduction section, conflict is defined as the impossibility to reach a goal that matters. In aeronautics, goals matter for safety reason such as avoiding loss of control and collisions with terrain or with an incoming aircraft. In our experimental scenario, we characterized conflict as unsolved when volunteers failed to manage the vertical separation with the incoming traffic according to aeronautical rule (the minimum vertical separation between two airplanes is 500 ft). We characterized conflict as solved when participants performed a relevant action that enabled them to recover from the critical situation (i.e. reducing selected speed with the dedicated FCU knob or manually taking over to keep on descending).

8.1.6 *Ocular Data Analysis*

Vertical and horizontal ocular movements were measured with the EyeTechSensor from Pertech and streamed in degrees relatively to the participants' head.

The ESIA was used for detecting the ocular events of three possible types (blinks, saccades and fixations). Though a temporal classification of fixations and saccades was suggested with the ESIA, this was not considered in the present work. Only the spatial classification was used, gaze moving under $30°/s$ with a dispersion threshold of $1°$ were considered as fixations. The number of occurrences of these ocular events can be used to characterize the visual processing behaviours.

We decided to search for a particular ocular behaviour that would characterize the moment of conflict. To do so, we choose a 15-second window of basic piloting activity (basic autopilot supervision and interaction) and compared it to the equivalent 15-second window following the rise of the conflict. We ran non-parametric Wilcoxon tests on fixations and saccades of different lengths (multiples of 20 ms – the inter-sampling period) to determine what event length best characterizes the conflict occurrence. We selected then all significant events and merged adjacent length-values in the same class. We obtained thus three classes: saccades of 120–160 ms duration, short fixations of 80–120 ms duration and long fixations of 240-260 ms. According to the mentioned above literature, we argue that the short fixations are not long enough to extract the complex infor-mation displayed

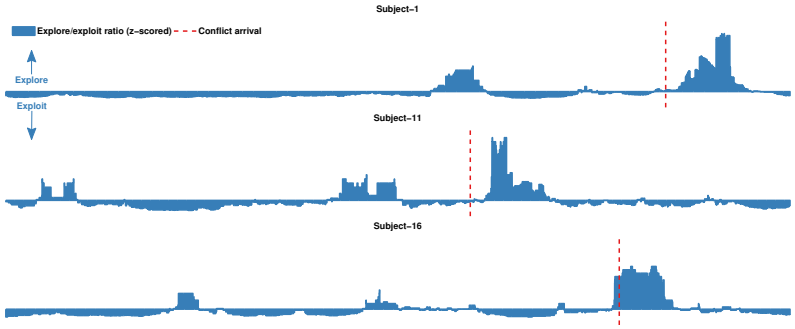


Figure 56: An example of explore/exploit ratio (z-scored) for three subjects. The red line shows the formal arrival of conflict.

in cockpit, and were therefore associated with infor-mation research. Keeping these three classes we express the Goldberg and Kotval's idea and compose a new adapted "explore/exploit" ratio defined as

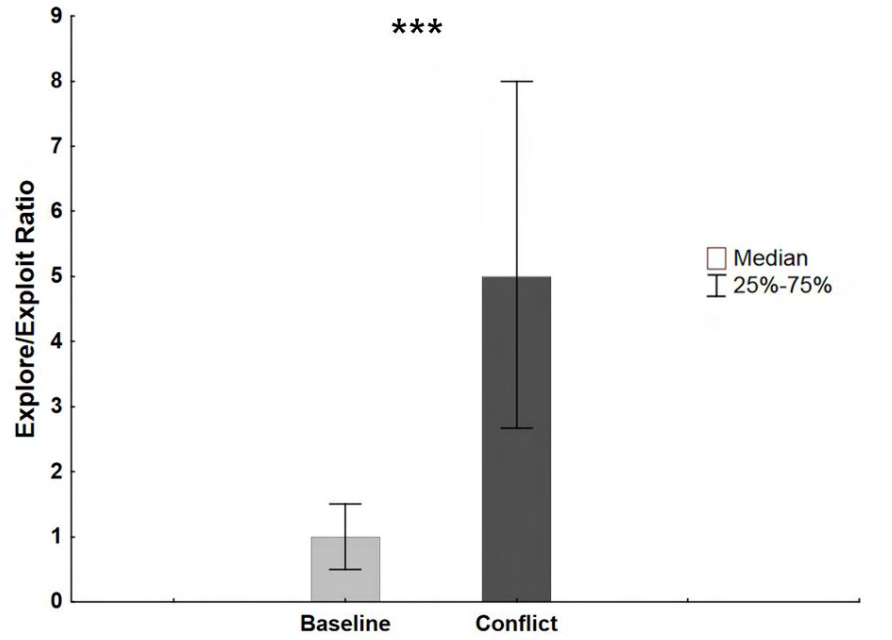
$$R = \frac{\text{saccades} + \text{short fixations}}{\text{long fixations}}.$$

On the one hand, when participants explore the interface, the number of saccades and short bottom-up fixations increases while the number of long top-down fixations decreases; leading to an increase of the ratio. On the other hand, when participants exploit the interface, long fixations dominate, and the ratio decreases. Figure 56 shows an example of the evolution of this defined ratio.

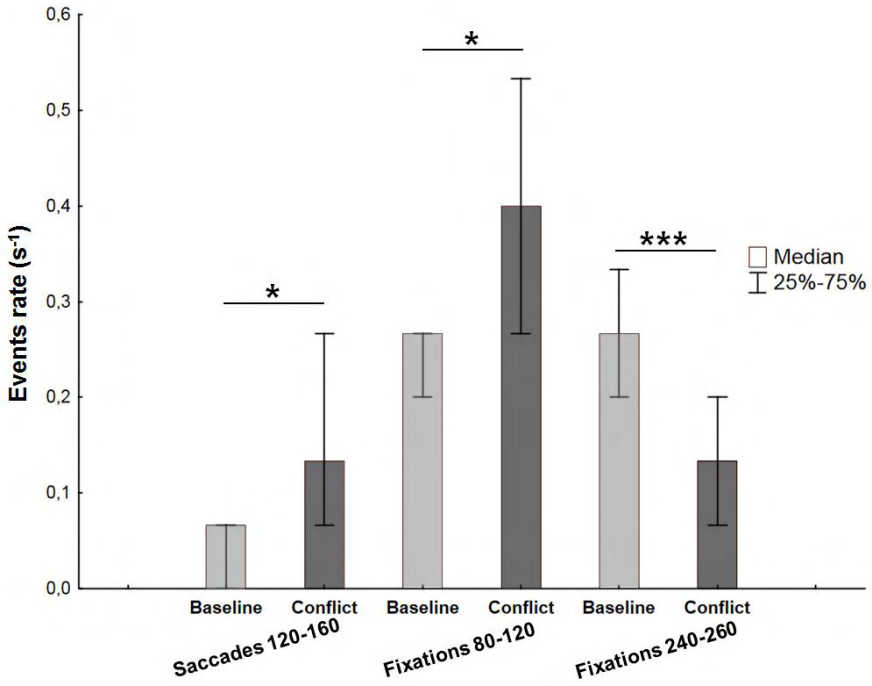
8.2 RESULTS

The debriefing session revealed that 14 out of 16 pilots have perceived the abnormal behavior of the autopilot. Among them, we found that only 7 out of 16 pilots understood the situation and initiated the correct recovery procedure within the critical time window of 30 s to avoid the incoming traffic (mean reaction time 18 s, SD= 9.3).

Regarding the Wilcoxon Matched pairs tests over the eye events, we found significant effects of the conflict corresponding to an increase of the saccade rate ($p < 0.05$; Z=2.04) and short fixations ($p < 0.05$; Z=2.36) and a decrease in the long fixation rate ($p < 0.001$; Z=3.40); (see Fig. 57b). In addition, the "explore/exploit" ratio over the 16 pilots (Fig. 57a) revealed a significant effect of the conflict with higher ratio for the conflict situation comparing to the "baseline" ($p < 0.001$; Z=3.29) confirming that the conflict implies higher exploration activity.



(a) Groupe results (N=16) of the"Explore/exploit" ratio. The conflict situation leads to an increase in the ratio value.



(b) Groupe results (N=16) over the eye events rate.

Figure 57: Ocular data group results. *:$p < 0.05$; ***:$p < 0.001$

8.3 CONCLUSION

The analysis of ocular events revealed that the volunteers exhibited higher visual search (more short fixations and saccades) to the detriment of infor-mation processing (less fixations) during conflict in comparison to baseline. It is worth noting that the ratio explore/-exploit (i.e search *vs*. process) dramatically increased and was five time higher during conflict than baseline. Indeed, "automation surprise" led to an excessive but inefficient visual search that prevents pilots from extracting the relevant infor-mation (i.e. the speed indicator).Taken together, these results tend to show that occurrence of automation surprise event summon up attentional resources toward conflict solving. This is akin with previous studies showing the link between conflicts, attentional impairment (Dehais, Causse, Vachon and Tremblay, 2012) and persistence in erroneous course of action (Billings, 1997, Dehais, Causse and Tremblay, 2011b).

This research intended to characterize and diagnose pilots' cognitive impairment due to the occurrence of a pilot-autopilot conflict following a silent mode change of the autopilot. Indeed, such conflicts create automation surprises, and jeopardize flight safety. The considered ocular metrics and defined ratio demonstrated that such conflicts impair attentional abilities, leading to an excessive visual search and an inability to extract relevant infor-mation. As focus of attention is mainly top-down and goal-driven (Mack and Rock, 1998b), it was not surprising that the participants had erratic eye movement with few ocular fixations. As their mental models and scheme of actions were insufficient, their attention was inefficiently bottom-up-driven.

9

DISCUSSION

RÉSUMÉ EN FRANÇAIS

Le dernier chapitre fait tout d'abord la synthèse puis apporte un regard critique sur l'ensemble du travail détaillé dans ce mémoire. Un ensemble de perspectives est alors énoncé, en tant que pistes d'améliorations ou nouvelles propositions.

La problématique abordée dans cette thèse concerne le suivi psychophysiologique des opérateurs (pilotes, conducteurs, contrôleurs aériens par exemple) en temps réel pour la détection des altérations de l'attention. Les attentes motivant un tel travail sont multiples: il s'agit d'une part de mieux comprendre les mécanismes de traitement de l'information et de prise de décision des opérateurs pour améliorer le design de leur environnement de travail, mais également de permettre le déploiement de contre-mesures en temps réel qui seraient un recours supplémentaires en cas de situation accidentogène. Ce manuscrit apporte deux contributions à cette problématique dans le contexte théo-rique suivant : la vision n'etant pas à même d'encoder les information de tout l'environnement, il incombe à l'attention d'optimiser la sélection d'infor-mation pour maximiser la pertinence de l'information encodée. Plusieurs modèles expliquent comment, vers quel cible, et quand les changements d'objet de l'attention sont effectués au sein du cerveau et on remarque alors que deux phénomènes sont particulièrement dangereux du point de vue des facteurs humains, la tunnélisation attentionnelle et la désorientation attentionnelle. Le fait que le changements d'objet de l'attention s'accompagnent quasiment systématiquement par un changement de position des yeux justifie de mesurer les mouvements oculaires pour étudier l'attention à l'aide d'un oculomètre.

Dans ce cadre, notre première proposition consiste en l'automatisation de la détection de la tunnélisation attentionnelle, à partir de métriques oculaires connues. Il est tout d'abord vérifié que l'expérience permet bien de générer deux groupes TUN=1 et TUN=0, et que les métriques connues se comportent comme attendu. Puis un classifieur temps réel est proposé pour estimer le niveau de tunnélisation attentionnelle de l'opérateur à chaque seconde de sa mission. Les performances de classification sont meilleures qu'avec un SVN, ce qui montre la pertinence de l'approche. De plus les lois d'inférence floues traduites en langage naturelles sont cohérentes avec celles utilisées par les experts en facteurs humains pour la détection de la tunnélisation attentionnelle. Il est néanmoins remarqué que l'utilisation d'un

tel système dans les domaines d'intérêts nécessite une très grande fiabilité qui nécessiterait encore d'approfondir l'approche proposée. Notamment, la prise en compte d'infor-mation liée au déroulement de la mission (phases en cours, objectifs, état système) permettrait d'affiner la stratégie de détection pour éviter les fausses alarmes. Il serait par ailleurs possible de prendre plus de métriques psychophysiologiques en compte, en utilisant d'autres techniques de mesure comme l'EEG ou en extrayant d'autres métriques de l'oculomètre comme la dilatation pupillaire. Une limite de notre approche vient de l'identification des phases de tunnélisation attentionnelle avec une approche experte avec seulement deux niveaux, 0 ou 1. En demandant au participant d'effecteur une tâche secondaire pendant leur mission, il serait possible d'avoir une estimation continue des ressources disponibles pour la seconde tâche et ainsi déterminer le niveau de tunnélisation attentionnelle à chaque instant de la mission. En utilisant ces données dans la base d'apprentissage, le classifieur produit serait continu également et permettrait la prédiction à court terme des phases à risque, en plus de leur détection.

Les métriques utilisées en entrée du classifieur nécessitent la connaissance de l'interface et d'importantes ressources pour être calculées en temps réel, ce qui complique le déploiement d'une telle solution sur le terrain. Notre deuxième proposition est donc une méthode de calcul de métriques dites "directes" qui se calculent facilement en temps réel tout en étant informatives du point de vue de la tunnélisation attentionnelle ou de la désorientation attentionnelle. Mais les méthodes proposées dans la littérature ne permettent une analyse assez fiable pour notre application. Pour ces raisons un algorithme d'identification de l'activité oculaire est proposé, l'ESIA. Deux métriques directes s'appuyant sur l'ESIA sont alors envisagées pour l'estimation de l'état attentionnel des opérateurs, l'identification des fixations focales versus ambiant et le ratio exploitation versus exploration. Il est montré dans un premier temps que l'ESIA présente de meilleur résultat que les algorithmes classiques, notamment en terme de résistance au bruit de mesure. Dans un second temps, il est vérifié que l'ESIA est robuste aux mouvements de compensation de la tête omniprésents en condition réelles avec un eye-tracker mobile, mais aussi que les métriques produites permettent l'estimation de métriques AOI-based. Plusieurs voies d'améliorations sont alors envisagées. Dans cette version la détection des clignements s'appuie sur les données manquantes de l'oculomètre. Utiliser un algorithme spécifique de détection des clignements améliorerait grandement la qualité de détection de cet état et permettrait d'utiliser cette information pour déployer les algorithmes déjà connus de détection de fatigue et d'endormissement, notamment pour l'automobile. Aussi la comparaison de performances avec les algorithmes de référence souffre de l'absence de base de données de test. Construire une telle

base de données permettrait de mieux comprendre les différences de performance des algorithmes étudiés. Une première perspective serait de prendre en compte le diamètre pupillaire qui est généralement calculé par les oculomètres. Cette donnée est également pertinente du point de vue de l'étude des états attentionnels en temps réel d'après de nombreux travaux. Il serait également possible de pousser l'analyse de corrélation entre les métriques calculées avec l'ESIA et les métriques de référence AOI-based pour augmenter le pouvoir informatif des métriques directes. Pour identifier la nature des fixations, il est possible de croiser les données de l'ESIA avec un celle d'un électroencéphalogramme ce qui viendrait affiner la classification temporelle proposée. Sur le modèle focal versus ambiant, il serait également possible d'étudier de plus longues séquences d'état, au travers de modèles de Markov à état cachés. Cette approche pourrait permettre d'identifier des signatures de certains schéma cognitifs spécifiques, ce qui permettrait de connaître en temps réel si le pilote est en phase de "supervision", "diagnostique" ou "pilotage aux instruments" par exemple.

Pour finir, il faut mettre en perspective que les propositions de méthodes de détection et métriques de l'état attentionnel des opérateurs présenté ici s'appuie exclusivement sur l'oculomètrie. D'autres moyens de mesure existent et sont tout à fait pertinant, comme l'électroencéphalographie, l'électrocardiographie, la spectroscopie proche infrarouge fonctionnelle. Toutes ces méthodes font l'objet d'étude prometteuses au sein du laboratoire d'accueil de ce travail de thèse dont la ligne directrice est le calcul temps réel du vecteur d'état pilote via la fusion de données objectives hétérogènes.

ENGLISH VERSION

The purpose of this Ph.D. was to propose attentional metrics that would be usable, generic, and real-time for human factors concerns in highly dynamical context such as automotive, aeronautics or drone-supervision. One promising avenue is indeed to monitor the operator's attentional state in order to interact on line with the system and possibly recover from accident prone situations.

9.1 MAIN RESULTS

9.1.1 *Implementation of the ANFIS*

Eye-tracking was chosen for its ability to measure the ocular activity which was demonstrated as being tightly linked with attention. Its high spatial resolution makes it a suitable tool for on-line analysis. One challenge with this technique is to extract the meaningful infor-mation out of the large amount of data that it produces. In that sense, many ocular metrics have been proposed, among which the fixation-saccade sequence became the referencel. From this sequence, basic metrics could be computed but also scanpaths or focal versus ambient fixations. Other more specific metrics are directly associated with the use of the interface. AOI based metrics were proven efficient in modelling the operators' infor-mation taking on different set-ups, but also in providing attentionl related indicators such as the NBAOI or the SWR.

Our first proposal was to demonstrate that such metrics could be used for on-line detection of attentional impairments in a realistic context. A previous robotic experiment had been designed specifically to provoke attentional tunnelling, and psycho-physiological metrics were computed. The statistical analysis confirmed that these metrics were relevant attentional tunnelling markers. Our goal was therefore to propose an automated inference system that could classify the periods of attentional tunnelling autonomously. We proposed a machine learning approach to this classification task. Our results show that ANFIS is appropriate. The training dataset was better modelled with this technique than with SVN, a reference machine learning technique. It was furthermore successful in identifying the attentional state of the 4 participants in the testing group, which proved that the approach is generalizable and would work on new participants. This inference system is therefore suitable for our purpose of detecting on line degraded attentional states from psycho-physiological measurements.

9.1.2 *The ESIA*

In order to overcome the limitations of the interface independent metrics, we proposed the Eye State Identification Algorithm which identifies saccades, fixations, and blinks. Two additional classifications of small versus large saccades, and express, short, medium and long fixations were proposed. The definition of the ESIA insures that the non-overlapping principle is respected. Furthermore, it overcomes some limitations of the current algorithms such as the lack of control of the identification process. The ESIA was compared to two reference algorithms the I-DT and I-VT. It had a better performance on our test dataset, and was robust to noise. Another interesting aspect of the ESIA is its identification strategy and additional classification that are explicit and which thresholds can adapted to fine tune the analysis. Subsequent basic and more complex ocular metrics computation was also presented.

The state distribution is a new metric which catches the eye activity as a limited resource. Our interpretation is that the state distribution can be seen as the result of the optimization of the eye resource. As such, changes in the distribution give a sense of the constraints (or cost functions) above the optimization process. This approach is supported by the fact that ESIA is suitable for identifying the ocular events seamlessly from the eye-in-head or the eye-in-the environment. Considering the eye-in-head permitted to observe how the ocular plants was piloted by the visual attention. We studied the correlation between the eye states in the distribution to elicit the ocular strategies. The correlation were compatible with the two modes of the visual system. This suggested to derive the exploit/explore ratio from the state distribution. Three definitions of the ratio were proposed according to Goldberg definition, the correlation analysis, and the ESIA temporal classification.

9.1.3 *Use cases of the ESIA and associated ratio*

In the robotic use case, the ability of the ESIA to produce relevant attentional metrics was tested. In the first part of the analysis of the ocular data, we considered all the fixations and saccades and investigated the possibility of classifying them as focal versus ambient. Two clusters were elicited from the fixation-saccade couples population. Differences were elicited in the extent of the saccades, but not in the mean fixation duration. Interestingly, the "focal" cluster had a centroid of few degrees that is consistent with previous litterature, but the centroid of the "ambient" cluster had an extent value three times greater than what was found by Follet and colleagues. One can explain this difference by the difference of the task: supervising the robot involved large extent saccades between the areas of interest of

the interface, whereas free-viewing of the picture of visual scene resulted in smaller extent saccades inside the visual scene.

The profiles of the saccade extent as a function of fixation duration did not elicit the classical relation used for classifying focal versus ambient fixations. One possible explanation is that supervising the robot involved continuous control over a long period in a environment that was dynamic, whereas free-viewing paradigm involved the repetition of short presentation of new images whose content is unexpected. The main difference at stake would therefore be the active top-down control while supervising the robot against the more relaxed bottom-up approach while free-viewing a visual scene. This difference of control have a dramatic impact on the ocular activity.

The exploit/explore ratio was then tested on three use-cases. On the first two, it was successful in identifying over focalization of the attention in the context of a robotic experiment and in the cockpit. In the BEA experiment the ratio saved time when dealing with the long records. The periods of interest were indeed directly identified as the values of the ratio above 2 standard deviation. The third use-case demonstrated the possibility of the ratio to detect "attentional disorientation" or "attentional lability" following an automation surprise. The ratio is therefore able to detect two opposite extreme attentional impairments: attentional tunnelling (over exploitation) and "attentional disorientation" (over exploration). It achieved our objective of providing an usable on-line interface independent attentional metrics.

Interface dependent metric such as the switching rate are very valuable for their ability to catch the interaction between the operator and the system. Studying the velocity distribution of all saccades versus transition saccades, it appeared that the velocity threshold chosen for separating small and large saccades would also likely separate transition saccades from the whole saccade population. A strong correlation was indeed found between the switching rate and the rate of large saccades. This highlights how independent metrics such as large saccades computed from the ESIA can be used to estimate interface dependent metrics such as the switching rate.

9.2 PERSPECTIVES

9.2.1 *ANFIS*

The ANFIS is a promising avenue for detecting degraded attentional states, but further work is required for integrating such classifier in systems where safety is at stake. Indeed, one direct application of such work would be to trigger adaptation of the system while in operation. Triggering such adaptation can be detrimental if not on purpose. The reliability of the classification should therefore be im-

proved. Adding new metrics to the classifier, such as the pupil dilation, is one interesting option. Another way to avoid spurious interventions would be to integrate infor-mation from the context itself. For instance, following the state of the system could allow to identify the periods of risky operations and restrict the detection of degraded attentional states on these periods.

Another limitation of our approach is the identification of the period of attentional tunnelling on the robotic experiment. First of all, we considered only a binary level of attentional tunnelling. But there is probably different levels of attentional tunnelling. As a consequence, the intermediate values providing by the classifier cannot be interpreted. Furthermore, there was only one attentional tunnelling probe in our experiment, the battery alarm. A great opportunity for improvement would be to use a double tasks paradigm that would test frequently the presence of attentional tunnelling on the secondary task. The performance indicators on the secondary task (response time, timing) could therefore be used as a continuous measurement of the level of attentional tunneling.

9.2.2 *ESIA*

The ESIA does not prevent from the threshold effects. Fuzzy domains would certainly be of interest for a limiting such effects. Of course, eye states are crisp values, and being one half-fixation and one half-saccade does not make sense. But Fuzzy logic could be used in the “black box" performing the identification, before defuzzyfication into crisp values corresponding to the eye states. Finally, machine learning techniques would certainly be successful for training fixation and saccade classifier with very high performance, without requiring the expert choice of dispersion and velocity threshold as it is the case with the ESIA.

Though the ESIA integrates blinks thanks to the status input, no specific algorithms were designed for detecting blinks from the raw eye-tracking data. The absence of a well controlled blink detection algorithm is one weakness of the ESIA in its current version, which threaten the integrity of the analysis. Blink detection is generally specific to the model of eye-tracker. For this work, blinks were computed directly by our eye-tracker as the samples on which the eye-tracker could not detect the pupil. But better blink detection algorithms could be integrated with the ESIA as another additional classification module that could run in parallel of the ESIA and feed the Status input. This would open the possibility for the ESIA to be used for studying fatigue of arousal from the blink activity. In its current version the log already provides the right infor-mation for blink characterization (blink duration, and peak velocity).

Though we analysed the link between the ESIA and interface dependent metrics such as the SWR, we would like to further investigate these links. A very high quality database recorded at 1000Hz of trainee pilots performing a multitasking task on a computer screen with AOIs was made accessible but was not exploited enough to be presented in this thesis. Finding correlation between interface dependent and interface independent would allow for extracting more infor-mation from the eye-in-head sequence. This would be of great interest in real plane cockpits where luminosity changes are of such great extents that the eye-in-head sequence is generally the only measurement accessible with using head-mounted systems.

Another ocular data that is not exploited in the ESIA which is an relevant attentional and cognitive indicator in the pupil diameter. The analysis of the pupil is currently under investigation by our research team, especially the correlation of the ambient luminosity and the pupil oscillations. Fast pupil oscillations are indeed a promising avenue for the study of cognitive activity but require to be decorrelated for the light adaptation oscillation. Another interesting approach also explored currently is the event related pupil dilation which could be combined with the ESIA for forming the Eye Fixation Related Pupil Dilatation (EFRPD), in which the pupil diameter would be used to differentiate fixations types, as EEG in the Eye Fixation Related Potential (EFRP).

The benchmark was limited to a relatively small dataset than in which saccades and fixation were manually coded. But this dataset has its own specificity that could influence the outcome of the benchmark. One very interesting perspective would be to provide a reference dataset to the community so the identification algorithms could be properly benchmarked. This seems to be a necessity for standardizing the definition of the identification algorithms and help the community compare works from different sources.

9.2.3 *Derived metrics from the ESIA*

We proposed to apply the zscore normalization on the ratio for limiting the impact of not adapting the thresholds to each participant. This indeed allowed for inter individual comparison, but in relative value only. The absolute value of the ratio is indeed lost by this operation. This absolute value has a physical meaning, it is how many times exploitation is more performed than exploration. The perspective of better controlling the ESIA analysis for each participant could result in a reliable absolute value of the ratio allowing for more precise comparisons. On main concern when using the zscore, is the choice of the reference window used for calculating the mean and the standard deviation. In our case, we used the overall mission. But

this choice would not be possible for an on-line analysis. The choice of the window should therefore be investigated.

Another promising approach for computing new metrics from the ESIA would be to use Hidden Markov Model (HMM). This approach was successful in detecting off-nominal ocular activity using the AOI sequence. We would like to adapt it to the ESV. HMM would extract the statistical profile of the transitions between samples state in the ESV. Several HMM could be fitted to different types of ocular activity on controlled dataset, and then used to detect these "signatures" in the ocular activity via the Vitterbi algorithm that gives the compatibility between the current sequence and the each HMM statistical model.

9.2.4 *Application*

This work contributed mainly on improving the eye-tracking analysis, and using eye-tracking together with heart rate gave good results with ANFIS. Multi-modality is a very promising perspective on the study of attention. Taken separately, many instruments propose a partial modelling of the attention. Experiments are carried out in our laboratory using functional Near Infra Red Spectroscopy (fNIRS), ElectroEncephaloGraphy (EEG) and ElectroCardioGraphy. Taken together, the classification of the operator state could be greatly improved. Our objectives is therefore to implement the ESIA and perform the subsequent analysis in real time, in a multi-modal environment. Our flight simulator is being constantly improved and more and more possibilities are offered for dynamical interfaces. We would like to implement a whole on line attentional impairment detection from psycho-physiological measurement together with countermeasures on the adaptive interfaces. Our goal is to aggregate the heterogeneous measurements and compute the operator state vector covering the main indicators on stress, fatigue, and attention.

10 APPENDIX : EYE-TRACKING DATA

In this section we focus on the data flow used in *head-mounted* systems only. The computation of the eye in the environment require several steps: pupil identification, calibration and mapping in the field camera, environment positioning in the field camera (which is equivalent to compute the head orientation in the environment).

PUPIL IDENTIFICATION

Eye cameras output images in greyscale (IR illumination). In the dark pupil paradigm, pupil should appear as the darkest pixels on the eye camera (except when wearing make-up, or in the presence of very dark eye lashes). An easy way to represent the pupil identification problem is to plot the spectrum of pixels grey level. If the eye camera is correctly set-up, then a peak appears at the low end of the spectrum. A simple threshold rejection will isolate the pixels that belong to the pupil.

These pixels are grouped in an ellipsoid. A simple ellipse fitting algorithm can be used for delimiting the pupil, or principal component analysis. In our eye-tracking system, a square is fitted around the group of pixel isolated. Using these techniques, the center of the pupil is found with its coordinates expressed in pixels relatively to the eye camera (see Figure 58).

PURKINJE REFLECTION DETECTION

As explained in the previous section, Purkinje corneal reflection (also called first Purkinje reflection) is often used as a reference point for the eye-tracker that measure the *vector* between the glints (IR dots due to reflection on the eye) and the pupil center. In these systems, the vector is the input instead of the pupil position in the eye camera.

CALIBRATION AND MAPPING

The next computational step consists of a mapping between the eye position in the eye camera and the gaze direction in the field camera.

The procedure is the following:

1. Participant and experimentalist agree on several noticeable targets in the environment. Participant is supposed to keep his

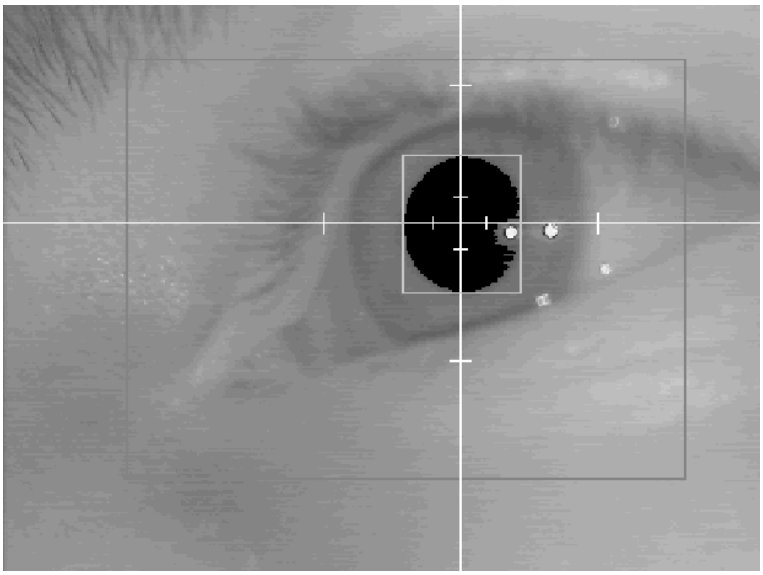


Figure 58: Pupil identification in dark pupil configuration using EyeTech-Sensor.

head still (this can be controlled) and only use his eye to go from a target to another at the experimentalist's will.

2. For each target, the target position on the *field* camera is inputed using the interface. Then the participant is asked to look at the target, and while the participant's head and eye are still, the eye position in the *eye* camera is recorded over a short period and averaged so as to be only one value.

3. A linear mapping is done between eye position in the *eye* camera and the target position in the *field* camera. When the mapping is done, each measured position of the eye will turned into a gaze direction in the field camera. At that stage, watching the field camera with the crosshair representing the gaze direction on top allow for an experimentalist to control where and what the participant is looking at (see Figure 59). But this analysis require the work of an expert and can not be automated.

4. The final step generally consists of converting the pixel coordinates in the field camera in degrees of eye rotation in the head. This conversion is quite simple, it only requires to know the angle of view of the camera. Approximations are considered to simplify the problem: the field camera is close enough to the eye to be considered as being at the same position (center of the camera is the center of the eye). Also, eye angles are rather small as the head tends to compensate for wide eye movements Duchowski (2007).

Figure 59: Gaze direction in the field camera, represented by the white crosshair, in the context of a flight simulation using EyeTechSensor

REAL-TIME DATA STREAMING

The calibration data are used once for all for computing the mapping. This mapping allows for the eye orientation to be computed in degrees directly. The EyeTechSensor streams each sample in real time in a tab-separated text line which contains 13 fields that are detailed in Table 8.

ENVIRONMENT REPOSITIONING

The next step is to input in the system the relevant object delimitation in a reference image. The world indeed appears in two dimensions from the field camera point of view. Relevant objects are represented by Areas Of Interest (AOI) (see Figure 60). But these AOI will not delimit the right instrument anymore if the head moves. That is the reason for repositioning the field camera image in the reference image.

The experimentalist therefore places specific IR markers in the visual scene, so that the position of the current image can be compared to the reference image where the objects positions are stored (see Figure 61). This finally allows for comparing the current gaze position with the objects position and therefore perform context related analysis.

1	Type	Current operation tag (calibration or recording)
2	Timestamp	Sample time in ms
3	TimeCode	Sample number (coded)
4	Eye X	Horizontal pupil position in pixels (eye camera)
5	Eye Y	Vertical pupil position in pixels (eye camera)
6	Field X	Horizontal eye orientation in pixels (field camera)
7	Field Y	Vertical Eye orientation in pixels (field camera)
8	Pup A	Not available at the time
9	Pup B	Not available at the time
10	Pup Surf	Not available at the time
11	Pup Diam	Not available at the time
12	Status	Binary code : 1 = ok, 0 = error
13	ERRCODE	Characterises the error if Status = 0

Table 8: One eye-in-head sample produced by the EyeTechSensor

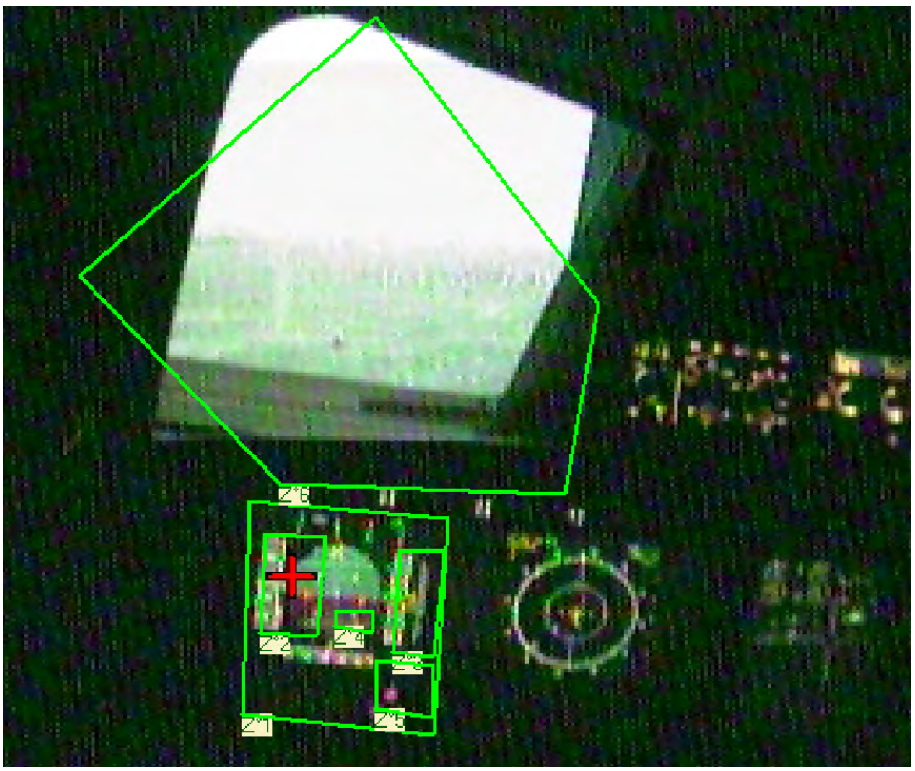


Figure 60: Areas of interest, in the context of a flight simulation. Each green polygon is an AOI. AOIs can overlap. EyeTechSensor

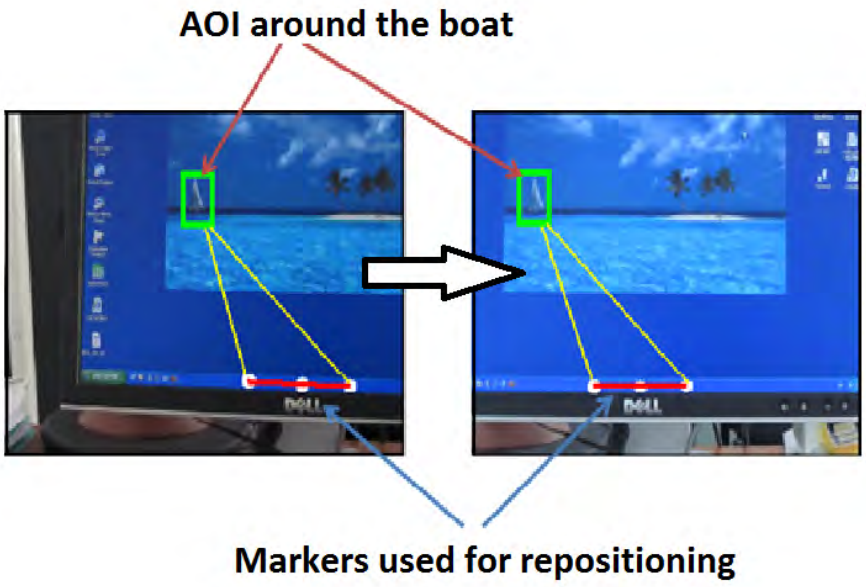


Figure 61: Environment Repositioning. Markers with specific geometry are automatically detected and used for repositioning the AOI around the object of interest.

Putting IR markers on the field camera, defining their geometrical characteristics in the video analyser, then perform the feature extraction on each frame of the video searching for the markers, computing the markers position and reposition the current image in the reference is an a posteriori process that is demanding in terms of time and processing power. Also, because the field camera is fixed relatively to the head, repositioning the environment in the field camera is finally equivalent to compute the head position in the environment. An alternative, that is as demanding as the previous option, is to place stylus on the participant's head. This stylus is positioned in space using IR video based recording thus providing the head position.

BIBLIOGRAPHY

Albert, W. S. and Liu, A. (2003). The effects of map orientation and landmarks on visual attention while using an in-vehicle navigation system, *Vision in Vehicles* **8**.

Aston-Jones, G. and Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance, *Annu. Rev. Neurosci.* **28**: 403–450.

Baccino, T. and Manunta, Y. (2005). Eye-fixation-related potentials: Insight into parafoveal processing, *Journal of Psychophysiology* **19**(3): 204–215.

Baddeley, A. (1972). Selective attention and performance in dangerous environements, *British journal of psychology* **63**(4): 537–546.

Bahill, A. T., Clark, M. R. and Stark, L. (1975). The main sequence, a tool for studying human eye movements, *Mathematical Biosciences* **24**(3): 191–204. 00630.

Bahrick, H. P., Fitts, P. M. and Rankin, R. E. (1952). Effect of incentives upon reactions to peripheral stimuli., *Journal of Experimental Psychology* **44**(6): 400.

Basic-T scan path (n.d.).

BEA (n.d.). Aeroplane state awareness during go-around, *Technical Report No. FRAN-2013-023*, Bureau Enquête Analyse.

Beck, V. M. and Luck, S. J. (2009). Temporal dynamics of the attentional template during visual search, *Target* **1**(2nd): 3rd.

Beigbeder, T., Coughlan, R., Lusher, C., Plunkett, J., Agu, E. and Claypool, M. (2004). The effects of loss and latency on user performance in unreal tournament 2003., *Proceedings of 3rd ACM SIGCOMM workshop on Network and system support for games*, pp. 144–151. Cited by 0218.

Bergasa, L., Nuevo, J., Sotelo, M., Barea, R. and Lopez, M. (2006). Real-time system for monitoring driver vigilance, *IEEE Transactions on Intelligent Transportation Systems* **7**: 63–77.

Billings, C. E. (1997). *Aviation automation: The search for a human-centered approach*, Lawrence Erlbaum Associates, Incorporated.

Boeing (2012). Statistical summary of commercial airplane accidents. worldwide operations 1959-2012, *Technical report*, Boeing Incorporation.

Bonner, M. A. and Wilson, G. F. (2002). Heart rate measures of flight test and evaluation, *The International Journal of Aviation Psychology* **12**(1): 63–77.

Broadbent, D. E. (1958). *Perception and Communication*, Oxford : Pergamon Press.

Brookings, J. B., Wilson, G. F. and Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control, *Biological Psychology* **42**(3): 361–377.

Brown, S. W. (1985). Time perception and attention: The effects of prospective versus retrospective paradigms and task demands on perceived duration, *Perception & Psychophysics* **38**(2): 115–124. Cited by 0263.
URL: *http://link.springer.com/article/10.3758/BF03198848*

Bruneau, D., Sasse, M. A. and McCarthy, J. D. (2002). The eyes never lie: The use of eye tracking data in HCI research, *Proceedings of the CHI*, Vol. 2, p. 25.

Cajal, S. R. and Azoulay, L. (1911). *Histologie Du Système Nerveux de l'Homme et Des Vertébrés*, A. Maloine.

Causse, M. (2010). *Influence de la récompense et de l'âge sur la performance de pilotage : une contribution de la neuroergonomie à la sécurité aérienne*, PhD thesis, Institut Supérieur de l'Aéronautique et de l'Espace, Toulouse.

Causse, M., Baracat, B., Pastor, J. and Dehais, F. (2011). Reward and uncertainty favor risky decision-making in pilots: evidence from cardiovascular and oculometric measurements, *Applied psychophysiology and biofeedback* **36**(4): 231–242. Cited by 0006.

Causse, M., Péran, P., Dehais, F., Caravasso, C. F., Zeffiro, T., Sabatini, U. and Pastor, J. (2013). Affective decision making under uncertainty during a plausible aviation task: An fMRI study., *NeuroImage* **71**: 19–29.

Chapman, P. R. and Underwood, G. (1998). Visual search of dynamic scenes: Event types and the role of experience in viewing driving situations, *Eye guidance in reading and scene perception* pp. 369–394.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears, *The Journal of the acoustical society of America* **25**(5): 975–979.

Codispoti, M., Bradley, M. and Lang, P. (2001). Affective reactions to briefly presented pictures, *Psychophysiology* **38**(3): 474–478.

Cohen, A. I. (1972). Rods and cones, *Physiology of Photoreceptor Organs*, Springer, pp. 63–110.

Cortes, C. and Vapnik, V. (1995). Support-vector networks, *Machine learning* **20**(3): 273–297.

Cowen, L., Ball, L. J. and Delin, J. (2002). An eye movement analysis of webpage usability, *Proc. People & Computers XVI: Memorable yet invisible* pp. 317–335.

Cox, E. (1992). Fuzzy fundamentals, *Spectrum, IEEE* **29**(10): 58–61.

Croner, L. J. and Kaplan, E. (1995). Receptive fields of P and M ganglion cells across the primate retina, *Vision research* **35**(1): 7–24.

Crundall, D., Underwood, G. and Chapman, P. (1999). Driving experience and the functional field of view, *Perception* **28**(9): 1075–1087.

De Oliveira, J. V. (1999). Semantic constraints for membership function optimization, *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on* **29**(1): 128–138.

De Valois, R. L. and De Valois, B. K. (1988). *Spatial vision*, Oxford University Press.

Dehais, F., Causse, M., Régis, N., Menant, E., Labedan, P., Vachon, F. and Tremblay, S. (2012). Missing critical auditory alarms in aeronautics: Evidence for inattentional deafness?, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 56, SAGE Publications, pp. 1639–1643.

Dehais, F., Causse, M. and Tremblay, S. (2011a). Mitigation of conflicts with automation, *Human Factors: The Journal of the Human Factors and Ergonomics Society* **53**(5): 448–460.

Dehais, F., Causse, M. and Tremblay, S. (2011b). Mitigation of conflicts with automation use of cognitive countermeasures, *Human Factors: The Journal of the Human Factors and Ergonomics Society* **53**(5): 448–460.

Dehais, F., Causse, M., Vachon, F., Régis, N., Menant, E. and Tremblay, S. (2014). Failure to detect critical auditory alerts in the cockpit evidence for inattentional deafness, *Human Factors: The Journal of the Human Factors and Ergonomics Society* **56**(4): 631–644. 00001.

Dehais, F., Causse, M., Vachon, F. and Tremblay, S. (2012). Cognitive conflict in human–automation interactions: a psychophysiological study, *Applied ergonomics* **43**(3): 588–595.

Dehais, F., Mercier, S. and Tessier, C. (2009). Conflicts in human operator–unmanned vehicles interactions, *Engineering Psychology and Cognitive Ergonomics*, pp. 498–507.

Dehais, F., Sisbot, E., Alami, R. and Causse, M. (2011). Physiological and subjective evaluation of a human-robot object hand-over task, *Applied Ergonomics* .

Dehais, F., Tessier, C. and Chaudron, L. (2003). GHOST: experimenting conflicts countermeasures in the pilot's activity, *IJCAI*, Vol. 3, pp. 164–168.

Dehais, F., Tessier, C., Christophe, L. and Reuzeau, F. (2010). The perseveration syndrome in the pilot's activity: guidelines and cognitive countermeasures, *Human Error, Safety and Systems Development*, pp. 68–80.

Deutsch, J. A. and Deutsch, D. (1963). Attention: some theoretical considerations., *Psychological review* **70**(1): 80.

Di Stasi, L. L., Antolí, A. and Cañas, J. J. (2011). Main sequence: An index for detecting mental workload variation in complex tasks, *Applied Ergonomics* **42**(6): 807–813.

Di Stasi, L. L., Catena, A., Cañas, J. J., Macknik, S. L. and Martinez-Conde, S. (2013). Saccadic velocity as an arousal index in naturalistic tasks, *Neuroscience & Biobehavioral Reviews* **37**(5): 968–975.

Di Stasi, L. L., Renner, R., Catena, A., Cañas, J. J., Velichkovsky, B. M. and Pannasch, S. (2012). Towards a driver fatigue test based on the saccadic main sequence: A partial validation by subjective report data, *Transportation research part C: emerging technologies* **21**(1): 122–133.

Di Stasi, L., Marchitto, M., Antolí, A. and Cañas, J. (2013). Saccadic peak velocity as an alternative index of operator attention: A short review, *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology* **63**(6): 335–343.

Ditchburn, R. W. and Ginsborg, B. L. (1953). Involuntary eye movements during fixation, *The Journal of physiology* **119**(1): 1.

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice*, Vol. 373, Springer.

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior., *Psychological review* **66**(3): 183.

Easterbrook, S. (1991). Handling conflict between domain descriptions with computer-supported negotiation, *Knowledge acquisition* **3**(3): 255–289.

Filipe, S. and Alexandre, L. A. (2013). From the human visual system to the computational models of visual attention: a survey, *Artificial Intelligence Review* .

Findlay, J. M. and Gilchrist, I. D. (2001). Visual attention: the active vision perspective, *Vision and attention*, Springer, pp. 83–103.

Findlay, J. M. and Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition, *Behavioral and Brain Sciences* **22**(04): 661–674.

Fitts, P. M., Jones, R. E. and Milton, J. L. (2005). Eye movements of aircraft pilots during instrument-landing approaches, *Ergonomics: Psychological mechanisms and models in ergonomics* **3**: 56.

Flemisch, F. O. (2000). Detecting usability problems with eye tracking in airborne battle management support, *Technical report*, DTIC Document.

Folkman, S. (1984). Personal control and stress and coping processes: a theoretical analysis., *Journal of personality and social psychology* **46**(4): 839.

Follet, B., Le Meur, O. and Baccino, T. (2011). New insights into ambient and focal visual fixations using an automatic classification algorithm, *i-Perception* **2**(6): 592–610.

Gibson, J. J. (1941). A critical review of the concept of set in contemporary experimental psychology., *Psychological Bulletin* **38**(9): 781.

Goldberg, J. H. and Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs, *International Journal of Industrial Ergonomics* **24**(6): 631–645.

Goldberg, J. H. and Wichansky, A. M. (2003). Eye tracking in usability evaluation: A practitioner's guide, *The mind's eye: Cognitive and applied aspects of eye movements research* pp. 493–516.

Goodale, M. A. and Milner, A. D. (1992). Separate visual pathways for perception and action, *Trends in neurosciences* **15**(1): 20–25.

Graf, W. and Krueger, H. (1989). Ergonomic evaluation of user-interfaces by means of eye-movement data, *Proceedings of the third International Conference on Human-Computer Interaction*, Vol. 1, Elsevier Science Inc., pp. 659–665.

Gregory, R. L. (1997). *Eye and brain: The psychology of seeing.*, Princeton university press.

Haddal, C. C. and Gertler, J. (2010). Homeland security: Unmanned aerial vehicles and border surveillance, DTIC Document.

Hastie, T., Tibshirani, R., Friedman, J. and Franklin, J. (2005). The elements of statistical learning: Data mining, inference and prediction, *The Mathematical Ingelligencer* **27**(2): 83–85.

Hauland, G. (2003). Measuring team situation awareness by means of eye movement data, *Proceedings of HCI International 2003*, Vol. 3, pp. 230–234.

Hayashi, M., Beutter, B. and McCann, R. S. (2005). Hidden markov model analysis for space shuttle crewmembers' scanning behavior, *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, Vol. 2, pp. 1615–1622.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H. and Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*, Oxford University Press.

http://www.qubitbiology.com (2010).
URL: *http://www.qubitbiology.com/wp-content/uploads/2010/07/S225-EOG-sensor.jpg*

Hubel, D. H. (1995). *Eye, brain, and vision.*, Scientific American Library/Scientific American Books.

Hubel, D. H. (2010). Eye, brain, and vision.
URL: *http://hubel.med.harvard.edu/book/bcontex.htm*

Inglis, M., Sutton, J. and McRandle, B. (2007). Human factors analysis of australian aviation accidents and comparison with the united states, *Technical Report B2004/0321*. ISBN 1 921092 59 9.
URL: *http://www.atsb.gov.au/media/29953/b20040321.pdf*

Irwin, D. E. (1992). Visual memory within and across fixations, *Eye movements and visual cognition*, Springer, pp. 146–165.

Jacob, R. and Karn, K. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises, *Mind* **2**(3): 4.

James, W. (1890). *The principles of psychology*, Harvard University Press.

Jang, K. E., Jeong, Y., Ye, J. C., Tak, S., Jung, J. and Jang, J. (2009). Wavelet minimum description length detrending for near-infrared spectroscopy, *Journal of biomedical optics* **14**(3): 034004–034004.

Just, M. A. and Carpenter, P. A. (1976). Eye fixations and cognitive processes* 1,* 2, *Cognitive Psychology* **8**(4): 441–480.

Kahneman, D. (1973). *Attention and effort*, Citeseer.

Karsh, R. and Breitenbach, F. (1983). Looking at looking: The amorphous fixation measure, *Eye movements and psychological functions: International views* pp. 53–64.

Kevin O'Regan, J., Deubel, H., Clark, J. J. and Rensink, R. A. (2000). Picture changes during blinks: Looking without seeing and seeing without looking, *Visual Cognition* **7**(1-3): 191–211.

Lai, D., Brandt, S., Luksch, H. and Wessel, R. (2011). Recurrent antitopographic inhibition mediates competitive stimulus selection in an attention network, *Journal of Neurophysiology* **105**(2): 793–805.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention., *Journal of Experimental Psychology: Human Perception and Performance* **21**(3): 451.

Le Meur, O. and Baccino, T. (2013). Methods for comparing scanpaths and saliency maps: strengths and weaknesses, *Behavior Research Methods* **45**(1): 251–266.

Lourens, P. F. (1989). Error analysis and applications in transportation systems, *Accident Analysis & Prevention* **21**(5): 419–426.

Macdonald, J. S. and Lavie, N. (2008). Load induced blindness., *Journal of Experimental Psychology: Human Perception and Performance* **34**(5): 1078.

Mack, A. and Rock, I. (1998a). *Inattentional blindness.*, The MIT Press.

Mack, A. and Rock, I. (1998b). *Inattentional blindness.*, The MIT Press.

Mandryk, R. and Atkins, M. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies, *International Journal of Human-Computer Studies* **65**(4): 329–347.

Marshall, S. (2002). The index of cognitive activity: measuring cognitive workload, *Human Factors and Power Plants, 2002. Proceedings of the 2002 IEEE 7th Human Factors Meeting*, pp. 7–5 to 7–9.

Marshall, S. P. (2000). Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity.

Martens, M. H. (2011). Change detection in traffic: Where do we look and what do we perceive?, *Transportation Research Part F: Traffic Psychology and Behaviour* **14**(3): 240–250.

Medical Atlas, e. (2012). Visual pathway page 7.
URL: *http://www.edoctoronline.com/medical-atlas.asp?c=4&id=21964&m=1&p=7&cid=1060&s=*

Mello-Thoms, C., Nodine, C. F. and Kundel, H. L. (2002). What attracts the eye to the location of missed and reported breast cancers?, *Proceedings of the 2002 symposium on Eye tracking research & applications*, ACM, pp. 111–117.

Meunier, F. (2009). On the automatic implementation of the eye involuntary reflexes measurements involved in the detection of human liveness and impaired faculties.

Michael, G. A., Garcia, S., Fernandez, D., Sellal, F. and Boucart, M. (2006). The ventral premotor cortex (vPM) and resistance to interference., *Behavioral neuroscience* **120**(2): 447.

Minassian, A., Granholm, E., Verney, S. and Perry, W. (2004). Pupillary dilation to simple vs. complex tasks and its relationship to thought disturbance in schizophrenia patients, *International Journal of Psychophysiology* **52**(1): 53–62.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions, *Quarterly journal of experimental psychology* **11**(1): 56–60.

Mumaw, R. J., Sarter, N. and Wickens, C. D. (2001). Analysis of pilots' monitoring and performance on an automated flight deck, *11th International Symposium on Aviation Psychology, Columbus, OH*.

Munn, S. M., Stefano, L. and Pelz, J. B. (2008). Fixation-identification in dynamic scenes: Comparing an automated algorithm to manual coding, *Proceedings of the 5th symposium on Applied perception in graphics and visualization*, ACM, pp. 33–42.

NASA-ARC (2012). Flight Deck Display Research Lab @ NASA Ames - Single Pilot Operations Technical Interchange Meeting.
URL: *http://humansystems.arc.nasa.gov/groups/FDDRL/SPO/agenda.php*

Newby, E. A. and Rock, I. (1998). Inattentional blindness as a function of proximity to the focus of attention, *Perception* **27**: 1025–1040.

Nyström, M. and Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data, *Behavior research methods* **42**(1): 188–204. 00083.

O'Hare, D., Wiggins, M., Batt, R. and Morrison, D. (1994). Cognitive failure analysis for aircraft accident investigation, *Ergonomics* **37**(11): 1855–1869.

Packwood, S., Hodgetts, H. M. and Tremblay, S. (2011). A multiperspective approach to the conceptualization of executive functions, *Journal of clinical and experimental neuropsychology* **33**(4): 456–470.

Pandey, B. (2013). Eyes facts – part II.

Parasuraman, R. and Rizzo, M. (2007). *Neuroergonomics: The brain at work*, Oxford University Press, USA.

Parasuraman, R. and Wickens, C. D. (2008). Humans: Still vital after all these years of automation, *Human Factors: The Journal of the Human Factors and Ergonomics Society* **50**(3): 511.

Parati, G., Saul, J., Di Rienzo, M. and Mancia, G. (1995). Spectral analysis of blood pressure and heart rate variability in evaluating cardiovascular regulation: a critical appraisal, *Hypertension* **25**(6): 1276–1286.

Pizziol, S., Dehais, F. and Tessier, C. (2011). Toward human operator "state" assessment, *1st ATACCS (Automation in Command and Control Systems)* .

Pizziol, S., Tessier, C. and Dehais, F. (2014). Petri net-based modelling of human–automation conflicts in aviation, *Ergonomics* **57**(3): 319–331.

Plainis, S., Murray, I. and Chauhan, K. (2001). Raised visual detection thresholds depend on the level of complexity of cognitive foveal loading, *Perception* **30**(10): 1203–1212.

Pomeranz, B., Macaulay, R. J., Caudill, M. A., Kutz, I., Adam, D., Gordon, D., Kilborn, K. M., Barger, A. C., Shannon, D. C. and Cohen, R. J. (1985). Assessment of autonomic function in humans by heart rate spectral analysis, *American Journal of Physiology-Heart and Circulatory Physiology* **248**(1): H151–H153.

Poole, A., Ball, L. J. and Phillips, P. (2005). In search of salience: A response-time and eye-movement analysis of bookmark recognition, *People and Computers XVIII—Design for Life* pp. 363–378.

Posner, M. (1980). Orienting of attention, *Quarterly Journal of Experimental Psychology* **32**(1): 3–25.

Posner, M. I. (1992). Attention as a cognitive and neural system, *Current directions in psychological science* **1**(1): 11–14.

Posner, M. I. and Fan, J. (2008). Attention as an organ system, *Topics in integrative neuroscience* pp. 31–61.

Posner, M. I. and Petersen, S. E. (1989). The attention system of the human brain, *Technical report*, DTIC Document.

Press, J., L. (2011). Sensorimotor dynamics and two visual systems: Shades of skeffington & brock part 1 | the VisionHelp blog.

Pêcher, C., Quaireau, C., Lemercier, C. and Cellier, J.-M. (2011). The effects of inattention on selective attention: How sadness and ruminations alter attention functions evaluated with the attention network test, *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology* **61**(1): 43–50.

Regis, N. (2011). Tunnélisation attentionnelle : Définitions de métriques physiologiques et comportementales pour diagnostiquer la tunnélisation attentionnelle chez un opérateur humain.

Regis, N., Dehais, F., Rachelson, E., Thooris, C., Pizziol, S., Causse, M. and Tessier, C. (2014). Formal detection of attentional tunnelling in human operator–automation interactions, *IEEE Transactions on Human-Machine Systems* pp. 1–11.

Regis, N., Dehais, F., Tessier, C. and Gagnon, J.-F. (2012). Ocular metrics for detecting attentional tunnelling, *Human Factors and Ergonomics Society–Chapter Europe, Toulouse, France* .

Rensink, R. A., O'Regan, J. and Clark, J. C. (1997). TO SEE OR NOT TO SEE: The need for attention to perceive changes in scenes, *Psychological science* **5**(8): 368—373.

Riid, A., Isotamm, R. and Rüstern, E. (2001). Transparency analysis of first-order takagi-sugeno systems, *Proc. 10th International Symposium on System-Modeling-Control*, pp. 165–170.
URL: *http://www.dcc.ttu.ee/Andri/teosed/trants.pdf*

Riid, A. and Rüstern, E. (2000). Transparent fuzzy systems and modeling with transparency protection, pp. 229–234.

RäMä, P. and Baccino, T. (2010). Eye fixation–related potentials (EFRPs) during object identification, *Visual Neuroscience* pp. 1–6.

Salvucci, D. and Goldberg, J. (2000). Identifying fixations and saccades in eye-tracking protocols, *Proceedings of the 2000 symposium on Eye tracking research & applications*, ACM, pp. 71–78.

Shannon, C. E. (1949). Communication in the presence of noise, *Proceedings of the IRE* **37**(1): 10–21.

Shappell, S. A. and Wiegman, D. A. (2003). A human error analysis of general aviation controlled flight into terrain accidents occurring between 1990-1998, *Technical report*, DTIC Document.

Shic, F., Scassellati, B. and Chawarska, K. (2008). The incomplete fixation measure, ACM, pp. 111–114.

Simons, D. J. and Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events, *Perception* **28**: 1059–1074.

Simons, D. J. and Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction, *Psychonomic Bulletin & Review* **5**(4): 644–649.

Simons, D. J. and Rensink, R. A. (2005). Change blindness: Past, present, and future, *Trends in cognitive sciences* **9**(1): 16–20.

Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression, *Statistics and computing* **14**(3): 199–222.

St. John, M., Kobus, D. A., Morrison, J. G. and Schmorrow, D. (2004). Overview of the darpa augmented cognition technical integration experiment, *International Journal of Human-Computer Interaction* **17**(2): 131–149.

Sugeno, M. and Kang, G. (1988). Structure identification of fuzzy model, *Fuzzy sets and systems* **28**(1): 15–33.

Sutter, A. (2003). Loyola university chicago - psyc/biol 240 - feature integration theory. 00000.

Svensson, E., Angelborg-Thanderz, M., Sjoberg, L. and Olsson, S. (1997). infor-mation complexity-mental workload and performance 1 in combat aircraft, *Ergonomics* **40**(3): 362–380.

Tatler, B. W., Gilchrist, I. D. and Land, M. F. (2005). Visual memory for objects in natural scenes: From fixations to object files, *The Quarterly Journal of Experimental Psychology Section A* **58**(5): 931–960.

Theeuwes, J. (1994). Stimulus-driven capture and attentional set: selective search for color and visual abrupt onsets., *Journal of Experimental Psychology: Human perception and performance* **20**(4): 799. 00484.

Thomas, L. and Wickens, C. (2004). Eye-tracking and individual differences in off-normal event detection when flying with a synthetic vision system display, *Human Factors and Ergonomics Society Annual Meeting Proceedings*, Vol. 48, pp. 223–227.

Tracy, J. I., Mohamed, F., Faro, S., Tiver, R., Pinus, A., Bloomer, C., Pyrros, A. and Harvan, J. (2000). The effect of autonomic arousal on attentional focus, *Neuroreport* **11**(18): 4037–4042.

Treisman, A. M. (1960). Contextual cues in selective listening, *Quarterly Journal of Experimental Psychology* **12**(4): 242–248.

Treisman, A. M. and Gelade, G. (1980). A feature-integration theory of attention, *Cognitive psychology* **12**(1): 97–136.

Tsai, Y., Viirre, E., Strychacz, C., Chase, B. and Jung, T. (2007). Task performance and eye activity: predicting behavior relating to cognitive workload, *Aviation, space, and environmental medicine* **78**(Supplement 1): B176–B185.

Unema, P. J. A., Pannasch, S., Joos, M. and Velichkovsky, B. M. (2005). Time course of infor-mation processing during scene perception: The relationship between saccade amplitude and fixation duration, *Visual Cognition* **12**(3): 473–494.

Vapnik, V. (2000). *The nature of statistical learning theory*, Springer-Verlag New York Inc.

Vapnik, V. N. (1999). An overview of statistical learning theory, *IEEE Transactions on Neural Networks* **10**(5): 988–999.

Velichkovsky, B. M., Dornhoefer, S. M., Pannasch, S. and Unema, P. J. (2000). Visual fixations and level of attentional processing, ACM, pp. 79–85.

Velichkovsky, B. M., Joos, M., Helmert, J. R. and Pannasch, S. (2005). Two visual systems and their eye movements: Evidence from static and dynamic scene perception, *Proceedings of the XXVII conference of the cognitive science society*, Lawrence Erlbaum Mahwah, NJ, p. 2283–2288.

Velichkovsky, B., Rothert, A., Kopf, M., Dornhöfer, S. and Joos, M. (2002). Towards an express-diagnostics for level of processing and hazard perception, *Transportation Research Part F: Traffic Psychology and Behaviour* **5**(2): 145–156.

Veltman, J. A. and Gaillard, A. W. K. (1996). Physiological indices of workload in a simulated flight task, *Biological psychology* **42**(3): 323–342.

Von Helmholtz, H. and Southall, J. P. C. (1925). *Helmholtz's treatise on physiological optics*, Optical Society of America.

Waraich, Q. R., Mazzuchi, T. A., Sarkani, S. and Rico, D. F. (2013). Minimizing human factors mishaps in unmanned aircraft systems, *Ergonomics in Design: The Quarterly of Human Factors Applications* **21**(1): 25–32.

Warner, R. M. and Strowman, S. R. (1995). Cardiovascular reactivity and positive/negative affect during conversations, *Journal of behavioral medicine* **18**(2): 141–159. Cited by 0025.

Weibel, R. E. and Hansman, R. J. (2006). Safety considerations for operation of unmanned aerial vehicles in the national airspace system. 00069.

Weltman, G., Egstrom, G. H. and Christianson, R. A. (1966). A system for underwater ergometry, *AIAA/USN 2nd Marine Systems & ASW Conference*.

Wickens, C. D. (2005). Attentional tunnelling and task management, *13th International Symposium on Aviation Psychology, Dayton, OH, USA*.

Wickens, C. D. (2008). Multiple resources and mental workload, *Human Factors: The Journal of the Human Factors and Ergonomics Society* **50**(3): 449–455.

Wickens, C. D. (2009). The psychology of aviation surprise: an 8 year update regarding the noticing of black swans, *ISAP Keynote* .

Wickens, C. D. and Alexander, A. L. (2009). Attentional tunneling and task management in synthetic vision displays, *International Journal of Aviation Psychology* **19**(2): 182–199.

Wickens, C. D., Hollands, G., J., Banbury, S. and Parasuraman, R. (2012). *Engineering Pshychology & Human Performance (4th Edition)*, pearson edn.

Wickens, C. D. and McCarley, J. S. (2007). *Applied attention theory*, Taylor & Francis.

Williams, L. (1995a). Peripheral target recognition and visual field narrowing in aviators and nonaviators, *The International Journal of Aviation Psychology* **5**(2): 215–232.

Williams, L. J. (1985). Tunnel vision induced by a foveal load manipulation, *Human Factors: The Journal of the Human Factors and Ergonomics Society* **27**(2): 221–227.

Williams, L. J. (1995b). Visual field tunnelling in aviators induced by memory demands., *Journal of General Psychology* **122**(225).

Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures., *International Journal of Aviation Psychology* .

Wilson, G. and Russell, C. (1999). Operator functional state classification using neural networks with combined physiological and performance features, *Human Factors and Ergonomics Society Annual Meeting Proceedings*, Vol. 43, pp. 1099–1102.

Yager, R. and Filev, D. (1994). Generation of fuzzy rules by mountain clustering, *Journal of Intelligent and Fuzzy Systems* **2**(3): 209–219.

Yarbus, A. L., Haigh, B. and Rigss, L. A. (1967). *Eye movements and vision*, Vol. 2, Plenum press New York.

Yeh, M., Wickens, C. D. and Seagull, F. J. (1999). Target cuing in visual search: the effects of conformality and display location on the allocation of visual attention, *Human Factors: The Journal of the Human Factors and Ergonomics Society* **41**(4): 524–542.

Zadeh, L. (1996). Fuzzy logic = computing with words, *IEEE Transactions on Fuzzy Systems* **4**(2): 103–111.

Zawidzki, T. and Bechtel, W. (2002). Gall's legacy revisited - decomposition and localization in cognitive neuroscience.

Zeki, S. (1993). A vision of the brain.